

Trabajo de Fin de Máster

## **Máster Universitario en Ingeniería Industrial**

### **Análisis y aplicación de diversos métodos y técnicas de previsión de ventas**

#### **MEMORIA**

**Autor:** Sofía Cámara Roldán  
**Director:** Gema Calleja Sanz  
**Convocatoria:** Otoño 2019-2020



Escola Tècnica Superior  
d'Enginyeria Industrial de Barcelona





## Resumen

El objetivo de este proyecto es describir una metodología de previsión de ventas que permita establecer un modelo de previsión adecuado dada una casuística concreta. Por ello, se realizan diversos análisis con el fin de conocer el impacto de las previsiones de ventas y qué técnicas de previsión son las más adecuadas para cada caso particular. La metodología descrita y los resultados de los análisis son aplicados a un caso práctico en el que se desea predecir las ventas de un establecimiento minorista de la compañía Walmart.

La metodología de previsión de ventas presentada considera diversos métodos de previsión. Con el fin de conocer qué métodos son los más adecuados para cada casuística, es importante conocer y entender el impacto de las previsiones de venta en la gestión de la cadena de suministro. Por ello, se analiza, considerando diferentes tipologías de producto, el impacto que las previsiones generan en los niveles de existencias y en la gestión de las actividades de todos los actores que forman la cadena de abastecimiento. Con el fin de mejorar la eficiencia de la cadena de suministro, se concluye que es importante predecir con una precisión elevada y en todas las etapas de la cadena, las ventas de productos que generan un gran número de ventas y/o ingresos, que tienen un elevado coste de stock, un alto riesgo a generar stock obsoleto o un tiempo de abastecimiento alto.

Con el fin de conocer que métodos de previsión deben ser considerados al aplicar la metodología descrita en un caso particular, se realiza un análisis de los diversos métodos y técnicas de previsión de ventas existentes. De este análisis se destaca que las ventas pueden ser previstas mediante métodos cualitativos o mediante métodos cuantitativos. Los segundos son en general más precisos y deben ser considerados cuando se dispone de un histórico de datos. Los métodos cuantitativos se dividen en métodos estadísticos basados en el análisis de series de tiempo y en métodos causales. Tradicionalmente las ventas son previstas utilizando modelos de series temporales; entre los que destacan las variaciones de los modelos autorregresivos integrados de medias móviles (ARIMA) y de los modelos de alisado exponencial. A pesar de ello, en el presente trabajo, se analizan también métodos causales de aprendizaje automático los cuales permiten, en general, la obtención de resultados más precisos.

En la resolución del caso práctico, donde se aplica la metodología presentada y se analiza la precisión de las previsiones aplicando diferentes modelos de previsión, se define un modelo de aprendizaje automático para realizar las previsiones de venta del establecimiento Walmart considerado. El modelo propuesto se basa en el aprendizaje combinado y considera las previsiones de un modelo de bosques aleatorios, en un 70%, y las de un modelo de regresión lineal, en un 30%; y, presenta una precisión del 96,57 %. En el caso práctico, se valora también el impacto de aplicar el modelo de ventas propuesto en el establecimiento Walmart.



# Sumario

<b>RESUMEN</b>	<b>3</b>
<b>SUMARIO</b>	<b>5</b>
<b>1. PREFACIO</b>	<b>9</b>
1.1. Origen del proyecto .....	9
<b>2. INTRODUCCIÓN</b>	<b>11</b>
2.1. Objetivos del proyecto .....	11
2.2. Alcance del proyecto .....	11
<b>3. LA IMPORTANCIA DE LA PREVISIÓN DE VENTAS EN LA GESTIÓN DE LA CADENA DE SUMINISTRO</b>	<b>13</b>
3.1. La cadena de suministro .....	13
3.1.1. Principales actores .....	13
3.1.2. Actividades .....	14
3.2. El efecto látigo en la cadena de suministro .....	15
3.3. El efecto de las previsiones de venta según la tipología de producto .....	16
3.4. Conclusiones parciales .....	17
<b>4. ANÁLISIS DE MÉTODOS Y TÉCNICAS DE PREVISIÓN DE VENTAS</b>	<b>19</b>
4.1. Consideraciones sobre las previsiones .....	19
4.2. Enfoques de la previsión .....	19
4.3. Modelos de previsión basados en el análisis de series temporales .....	21
4.3.1. Introducción a las series temporales .....	21
4.3.2. Modelo autorregresivo .....	24
4.3.3. Modelo de medias móviles .....	24
4.3.4. Modelo autorregresivo de medias móviles .....	24
4.3.5. Modelo autorregresivo integrado de medias móviles .....	25
4.3.6. Modelo estacional autorregresivo integrado de medias móviles .....	26
4.3.7. Modelo estacional autorregresivo integrado de medias móviles con entrada exógena .....	27
4.3.8. Modelo de alisado exponencial simple .....	27
4.3.9. Modelo de alisado exponencial doble .....	27
4.4. Modelos de previsión de causalidad basados en el aprendizaje automático .....	28
4.4.1. Introducción al aprendizaje automático .....	28

4.4.2.	Algoritmos de regresión .....	29
4.4.3.	Algoritmos basados en casos .....	31
4.4.4.	Algoritmos basados en árboles de decisión .....	32
4.4.5.	Algoritmos basados en redes neuronales .....	34
4.4.6.	Algoritmos de conjunto.....	35

## **5. METODOLOGÍA PARA LA SELECCIÓN Y APLICACIÓN DE UN MODELO DE PREVISIÓN DE VENTAS \_\_\_\_\_ 37**

5.1.	Precisión necesaria del modelo de previsión dependiendo de la tipología de producto .....	37
5.2.	Exploración y manipulación de los datos.....	38
5.3.	División de los datos para la valoración de la precisión del modelo .....	39
5.3.1.	División de los datos en datos de entrenamiento y de prueba .....	39
5.4.	Aplicación de modelos candidatos y valoración de los resultados .....	40
5.4.1.	Selección de un modelo candidato .....	40
5.4.2.	Selección de los mejores parámetros para el modelo candidato .....	40
5.4.3.	Entrenamiento y valoración del modelo candidato.....	41
5.5.	Selección del modelo definitivo.....	41
5.5.1.	Métodos combinados de aprendizaje .....	41
5.6.	Aplicación y actualización del modelo definitivo .....	42

## **6. APLICACIÓN Y VALORACIÓN DE LOS MÉTODOS DE VENTAS ANALIZADOS EN UN CASO PRÁCTICO \_\_\_\_\_ 43**

6.1.	Presentación del caso práctico .....	43
6.2.	Precisión necesaria del modelo de previsión dependiendo de la tipología de producto .....	45
6.3.	Exploración y manipulación de los datos.....	46
6.4.	División de los datos para la valoración de los modelos de previsión candidatos.....	50
6.5.	Aplicación de diversos modelos candidatos y valoración de los resultados	50
6.5.1.	Modelo simple de medias móviles .....	50
6.5.2.	Modelo estacional autorregresivo integrado de medias móviles con entrada exógena .....	51
6.5.3.	Modelo de regresión lineal .....	55
6.5.4.	Modelo de k-vecinos más cercanos.....	58
6.5.5.	Modelo de bosques aleatorios .....	61

6.5.6. Modelo de redes neuronales.....	63
6.5.7. Modelo XGBoost.....	66
6.6. Selección del modelo definitivo .....	69
6.6.1. Comparación de los modelos candidatos considerados .....	69
6.6.2. Selección del modelo definitivo en base al método de aprendizaje combinado .....	70
6.7. Aplicación y actualización del modelo definitivo.....	71
6.7.1. Ventajas de aplicar el modelo de previsión de ventas.....	72
6.7.2. Posibles barreras que pueden impedir la correcta implementación del modelo de previsión de ventas .....	74
<b>7. PLANIFICACIÓN TEMPORAL Y COSTES .....</b>	<b>75</b>
<b>8. IMPACTO AMBIENTAL .....</b>	<b>77</b>
<b>CONCLUSIONES .....</b>	<b>79</b>
<b>AGRADECIMIENTOS .....</b>	<b>83</b>
<b>BIBLIOGRAFÍA .....</b>	<b>85</b>





# 1. Prefacio

## 1.1. Origen del proyecto

La industria 4.0 es la revolución industrial actual y consiste en la digitalización y optimización de procesos gracias al uso de la inteligencia artificial y de infraestructuras inteligentes [1]. Son muchas las empresas que intentan adaptarse y beneficiarse de dicha revolución. A pesar de ello, son mayoría el número de compañías que a día de hoy todavía confían en procesos manuales y lejanos a los ofrecidos por la nueva revolución industrial.

Uno de los ámbitos donde la irrupción de la nueva revolución industrial permite la mejora de procesos y de su eficiencia es en la gestión de la cadena de suministro. El flujo de comunicación, y la calidad de este, entre los diferentes actores que conforman las cadenas de abastecimiento es primordial para la buena gestión de los procesos involucrados en el suministro de productos y servicios; y la industria 4.0 ofrece numerosas posibilidades para mejorar el flujo y la calidad de las comunicaciones.

Una de las bases de la industria 4.0 es el uso de la inteligencia artificial. Dicho concepto permite, entre otras cosas, predecir mediante el uso del aprendizaje automático el comportamiento de procesos y sistemas; y, en consecuencia, prepararse con anterioridad a acontecimientos futuros. Este concepto es, en muchos casos, aplicable de forma simple y puede conllevar grandes mejoras en procesos.

Una de las aplicaciones del aprendizaje automático, en el ámbito comentado de la cadena de suministro, es en la previsión de ventas. Una previsión adecuada permite comunicar y predecir de manera precisa la demanda de productos y/o servicios que solicitarán los diferentes actores de la cadena de suministro; permitiendo esto optimizar los procesos necesarios para el abastecimiento de dichos productos y/o servicios.

Las técnicas de aprendizaje automático permiten mejorar notablemente las predicciones de ventas y ello genera grandes beneficios en la gestión de la cadena de abastecimiento. Además, la incorporación de estas técnicas supone una baja inversión de recursos. A pesar de ello, la previsión de ventas sigue realizándose de manera manual y poco precisa en la mayoría de empresas, conllevando ello grandes ineficiencias en las cadenas de suministro.

El presente trabajo pretende solucionar el problema de la mala gestión de las previsiones de ventas en las empresas. Por ello, se desea presentar una metodología de previsión que permita a las compañías realizar una gestión eficiente de las previsiones de ventas y, en consecuencia, de todas las actividades de su cadena de abastecimiento. Dicha metodología

considerará modelos de técnicas de aprendizaje automático y permitirá, por tanto, la adopción de las tendencias de la industria 4.0 cuando esto sea oportuno.

## 2. Introducción

### 2.1. Objetivos del proyecto

El presente proyecto tiene como principales objetivos los siguientes puntos básicos:

- Valorar la importancia y las consecuencias de disponer de una previsión de ventas adecuada en la gestión de la cadena de suministro
- Analizar diversos métodos de previsión que pueden ser utilizados en la previsión de ventas diferenciándolos entre técnicas de análisis de series de tiempo y técnicas de aprendizaje automático
- Presentar una metodología para la aplicación y selección de un modelo de previsión adecuado dada una casuística concreta
- Aplicar la metodología presentada a un caso práctico considerando las diversas técnicas de previsión analizadas

En consecuencia, los principales capítulos de la memoria son los siguientes:

- Capítulo 3: presenta un análisis cualitativo de las consecuencias e importancia de las previsiones de venta en la gestión de la cadena de suministro.
- Capítulo 4: analiza técnicas de previsión que son aplicables a la previsión de ventas. En primer lugar, se analizan diversas técnicas basadas en el análisis de series de tiempo. A continuación, se analizan técnicas de aprendizaje automático agrupándolas por similitud.
- Capítulo 5: presenta una metodología para la obtención de un modelo de previsión adecuado dada una casuística concreta y basada en analizar diversas técnicas de previsión.
- Capítulo 6: presenta los resultados de analizar diversas de las técnicas presentadas, mediante la metodología mencionada, a un caso práctico. El caso práctico consiste en predecir las ventas de un establecimiento de la compañía Walmart.

### 2.2. Alcance del proyecto

El presente proyecto analiza diferentes técnicas de predicción aplicables a la previsión de

ventas y basadas en el análisis de series de tiempo y en el aprendizaje automático o *machine learning*, en inglés. El proyecto no pretende ser una guía de todas las técnicas de previsión de ventas existentes; si no un análisis de las técnicas más comunes y/o adecuadas para la previsión de ventas.

El proyecto presenta también los resultados de aplicar los modelos analizados y la metodología presentada a un caso práctico. El caso práctico se centra en predecir las ventas de un establecimiento de Walmart, corporación multinacional que opera cadenas de grandes almacenes de descuento. Dicho caso es la simplificación de una competición de Kaggle, comunidad online de científicos de datos, creada por Walmart con el fin de descubrir y contratar nuevos talentos. El caso práctico se trata de una simplificación de dicha competición con el objetivo de reducir el tiempo de computación a la hora de resolverlo. Por tanto, el proyecto no presenta una resolución a la competición propuesta por Walmart en Kaggle, únicamente se utilizan parte de los datos presentados en dicha competición.

### 3. La importancia de la previsión de ventas en la gestión de la cadena de suministro

#### 3.1. La cadena de suministro

La cadena de suministro o abastecimiento es el conjunto de individuos, organizaciones, recursos, actividades y tecnología involucrados en la creación y venta de un producto o servicio. La cadena de suministro está formada por diferentes actores que realizan diversas actividades y que deben mantener un flujo de información constante entre ellos con el fin de que todos los procesos involucrados sean lo más eficientes posibles.

Una mayor eficiencia en la cadena permite realizar sus procesos a menor precio y en menor tiempo. Por ello, las empresas proveedoras de productos y/o servicios invierten esfuerzos en realizar una gestión efectiva de la cadena de abastecimiento.

##### 3.1.1. Principales actores

Los principales actores que intervienen en la cadena de suministro son, tal y como se muestra en la Figura 3.1: los proveedores, los fabricantes, los distribuidores, los detallistas y los consumidores.

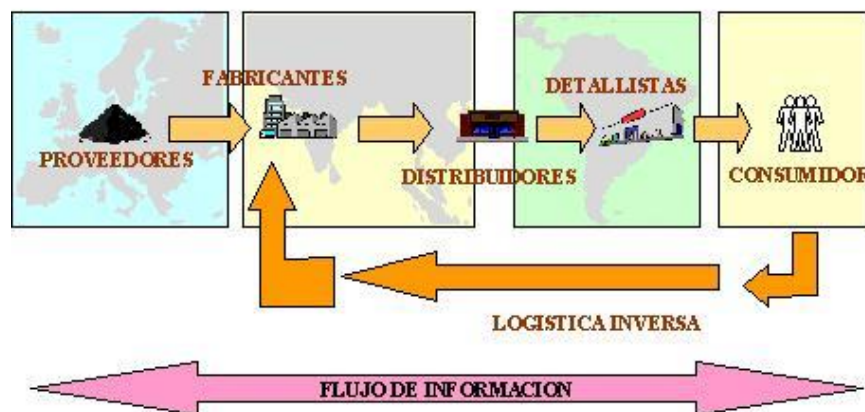


Figura 3.1 Principales actores de la cadena de suministro Fuente: [2]

Todos ellos conforman la cadena de abastecimiento de producto más habitual; que consiste en elaborar un producto final a partir de materias primas procedentes de diversos proveedores y distribuirlo por diversos canales al consumidor final. De todas formas, la cadena de suministro no siempre es la misma, ya que depende del producto o servicio que se esté

ofertando. Por tanto, no es necesario que todos los principales actores numerados estén siempre presentes y pueden existir actores adicionales.

Las funciones de los principales actores considerados se enumeran a continuación.

- Los proveedores son el elemento inicial de la cadena de suministro y abastecen, generalmente de manera periódica, de bienes, productos o servicios primarios que son necesarios para la obtención del bien o servicio que será ofertado al consumidor final.
- Los fabricantes permiten la elaboración de un producto final, que será el ofertado al consumidor, a partir de la materia prima procedente de los proveedores.
- Los distribuidores son las personas u organizaciones encargadas de la distribución de los productos finales a los diversos puntos de venta.
- Los detallistas son los comerciantes que disponen de los puntos de venta que ofertan el producto final al consumidor final.

### **3.1.2. Actividades**

Al igual que los actores, las actividades que se realizan en la cadena de suministro varían según el producto o servicio ofertado. A pesar de ello, existen algunas actividades que se pueden considerar necesarias en cualquier cadena de suministro de un producto para asegurar el buen funcionamiento de esta. Algunas de estas actividades son: planificación, gestión de existencias, procesamiento de órdenes de compra y traslados de productos.

La actividad que tiene una mayor repercusión en la eficiencia de la cadena de abastecimiento es la planificación [3]. Todos los actores de la cadena de suministro necesitan realizar una planificación de sus actividades internas para asegurar que ofertan a sus clientes los servicios y/o productos solicitados en el tiempo deseado. Para poder realizar una programación adecuada de sus actividades, todos los actores de la cadena necesitan estimar de forma precisa la cantidad de producto que su cliente solicitará.

Considerando la cadena de abastecimiento presentada en la Figura 3.1, los detallistas predicen las ventas que realizarán los consumidores finales en sus puntos de venta e informan a los distribuidores de las cantidades de producto que desean para poder satisfacer dichas ventas.

Los distribuidores, gracias a la información proporcionada por los detallistas y a sus propias

predicciones, planifican sus actividades internas para proporcionar a los detallistas los productos que prevén que demandarán y solicitan también los productos que necesiten para ello a los fabricantes. De forma análoga, los fabricantes solicitan las materias primas a los proveedores.

Tal y como es evidente, si los detallistas predicen de manera adecuada las ventas que realizarán gracias al consumidor final y no solicitan a los distribuidores cantidades de producto inesperadas con poca antelación, todos los actores de la cadena de suministro serán beneficiados; permitiéndoles a todos ellos predecir adecuadamente la demanda que su cliente realizará y por tanto planificar acordemente sus actividades internas.

### 3.2. El efecto látigo en la cadena de suministro

El efecto látigo o *bullwhip effect* es un fenómeno observado en la cadena de suministro que consiste en que fluctuaciones en la demanda esperada conllevan ineficiencias en la cadena de suministro asociadas a las variaciones de los niveles de existencias; teniendo estas ineficiencias un mayor impacto en los actores de la cadena más alejados del consumidor final [4]. El efecto apareció por primera vez en Jay Forrester's Industrial Dynamics (1961) y por ello es también conocido como efecto Forrester.

En la Figura 3.2 se puede observar el efecto látigo y la influencia que este provoca en los niveles de stock en los diferentes actores de la cadena de suministro. En concreto se representa el efecto que provoca que la cantidad demandada por el consumidor se incremente cuando se esperaba que esta fuese constante. Dicho incremento provoca que, a medio plazo, la cantidad demandada por el detallista al distribuidor aumente en mayor proporción y que aumente aún más la cantidad que el detallista solicita al fabricante. Tal y como puede observarse en la Figura 3.2, dichas variaciones en la demanda provocan fluctuaciones en los niveles de existencias de los diferentes actores. El aspecto de estas fluctuaciones recuerda a las ondulaciones que provoca agitar un látigo: las fluctuaciones son menores en el origen (consumidor final) y superiores al final (fabricante).

Se debe destacar que dicho efecto se produce si la demanda esperada del consumidor es distinta a la demanda real. El efecto látigo no se produce si la demanda del consumidor es diferente respecto a períodos anteriores pero la demanda se ha predicho correctamente por parte de los detallistas y se ha informado correctamente al resto de actores de la cadena de suministro, permitiéndoles a estos una buena predicción de sus respectivas demandas.



Figura 3.2 Representación del efecto látigo y su influencia en los niveles de stock en los diversos actores de la cadena de suministro Fuente: commons.wikipedia.org

### 3.3. El efecto de las previsiones de venta según la tipología de producto

Tal y como se ha presentado, una mala previsión de las ventas por parte de los detallistas, genera fluctuaciones en las existencias e ineficiencias en la gestión de las actividades de todos los actores de la cadena de suministro. A continuación, se evalúa el efecto de las previsiones de venta según la tipología de producto.

Por un lado, se pueden considerar diferentes tipos de producto considerando la rotación de ellos. Existen los denominados productos de alta rotación, los cuales tienen un ritmo elevado de entradas y salidas. Es decir, generan un número elevado de ventas y a la vez son repuestos con frecuencia. En contraposición, existen los productos de baja rotación, los cuales generan un bajo número de movimientos de entrada y salida. Respecto a la precisión de las previsiones de venta, se debe destacar que, si la precisión es baja, ello conllevará mayores consecuencias en los productos de alta rotación que en los de baja; el efecto látigo generará mayores fluctuaciones en los niveles de existencias y los actores verán sus beneficios afectados en mayor medida en el caso de los productos de alta rotación.

Por otro lado, se deben considerar los productos nuevos, que se incorporan por primera vez a la cartera de productos de una empresa. Estos productos son frecuentemente no considerados en las previsiones de venta, ya que la falta de conocimiento del producto dificulta



su previsión. De todas formas, se debe tener presente que se trate de un producto nuevo o no, una mala previsión de sus ventas generará los mismos efectos en la gestión de la cadena de suministro. Por ello, se debe destacar la importancia de la precisión de las previsiones de venta también para los productos nuevos; ya que en caso de que estos se conviertan desde un inicio en productos de alta rotación, una mala previsión conllevará pérdidas de beneficios e ineficiencias en la gestión.

Además, existen también otros factores que pueden afectar la necesidad de disponer de una previsión de ventas precisa como es el *lead time* o tiempo de espera para abastecerse de un producto, el precio de venta o el precio de almacenaje de un producto. Los productos de alta rotación se asocian a un *lead time* bajo y a los de baja rotación con un *lead time* alto. Aun así, pueden existir productos que generen un gran número de ventas, pero presenten un tiempo de abastecimiento alto y viceversa; por tanto, se debe tener presente que cuanto más elevado sea el *lead time*, más críticas serán las previsiones de ventas. Se debe considerar también tanto el precio de venta de un producto como el precio de almacenaje de este; ya que pueden existir productos de baja rotación que presenten elevados precios de venta o elevados costos de almacenamiento y que, por tanto, sus precisiones de venta sean críticas.

Por ello, se debe tener presente que debido a que pueden existir productos de alta rotación donde la precisión de las previsiones de venta no sea crítica y que sí lo sea para algunos productos de baja rotación; es necesario evaluar la criticidad de la precisión de las previsiones de venta de cada producto considerando todos los factores mencionados.

### 3.4. Conclusiones parciales

Tal y como se ha analizado en el presente apartado, la predicción de las ventas por parte de los detallistas es una de las actividades que repercuten en mayor medida a la eficiencia de la cadena de suministro.

Como se ha visto, una buena predicción de las ventas por parte de los detallistas, permite a los diferentes actores de la cadena de suministro realizar una gestión eficiente de sus actividades y mantener unos niveles de existencias estables.

Debido a la importancia e impacto de la predicción de la demanda del consumidor final por parte de los puntos de venta, el presente proyecto analiza técnicas de previsión de ventas con el fin de concluir cómo contribuir a aumentar la eficiencia de la cadena de suministro mejorando las predicciones de venta.



## **4. Análisis de métodos y técnicas de previsión de ventas**

En el apartado anterior se ha concluido que una previsión de ventas adecuada permite mejorar la eficiencia de la cadena de suministro asociada. Con el fin de conocer cómo predecir las ventas de forma efectiva, en el presente apartado se analizan diversos aspectos de las previsiones, aplicables a las previsiones de ventas, y se presentan diferentes métodos de previsión.

### **4.1. Consideraciones sobre las previsiones**

Las previsiones pretenden establecer comportamientos futuros de cierta variable o grupo de variables en base a información del pasado y conociendo, en algunos casos, acontecimientos futuros. Además, la información de la que se dispone, no siempre corresponde al objeto que se desea prever. En el caso de las previsiones de ventas, se utiliza información de ventas del pasado y se conocen ciertos acontecimientos futuros (acciones comerciales, apertura de nuevos establecimientos de venta, etc.) para predecir ventas futuras; cuando en realidad se desea conocer la demanda futura. Las ventas podrían ser inferiores a la demanda si no se es capaz de satisfacer la totalidad de esta.

Por todo ello se debe ser consciente de la complejidad de realizar previsiones exactas. De todas formas, el propósito de las previsiones no es que sean exactas sino útiles. Además de la precisión, se deben considerar otros aspectos como la inversión económica y el tiempo de cálculo necesario para obtenerlas. En general, a mayor precisión mayor inversión económica y mayor tiempo de cálculo. Dependiendo del objeto a predecir, la precisión que se considerará útil y la inversión económica y de tiempo de cálculo que se considerará aceptable variará.

### **4.2. Enfoques de la previsión**

Existen diferentes métodos y técnicas de previsión. Dependiendo del objeto a predecir y de la casuística concreta, el método y la técnica más adecuada variarán. En términos generales existen dos enfoques: utilizar métodos cualitativos o métodos cuantitativos [5]. Ambos enfoques son utilizados en el ámbito de previsión de ventas.

Se recurre a métodos cualitativos cuando la situación no es clara o cuando no existen suficientes datos pasados del objeto a predecir para establecer un comportamiento futuro fiable. Dichos métodos requieren de intuición y experiencia de expertos para su correcto diseño y aplicación. Dentro de los métodos cualitativos se destacan las técnicas de juicio y las

técnicas de contaje. Las primeras se centran en captar opiniones de una muestra de encuestados; entre estas técnicas se destacan el método Delphi y el taller de expertos. Por otro lado, las técnicas de contaje captan juicios de una muestra y proyectan las respuestas a un universo más amplio; en ellas se incluyen las investigaciones de mercado.

Como se ha comentado, se suele recurrir a estos métodos cuando no se dispone de datos, como es el caso de predecir las ventas de productos nuevos; o cuando la situación no es clara, por ejemplo, si se espera un cambio brusco en las temperaturas y se desea predecir las ventas de un producto estacional como ventiladores o aires acondicionados. De todas formas, en ambos casos se pueden utilizar métodos cuantitativos si se desea; ya que se puede estimar las ventas de un producto del que no se dispone histórico de datos analizando un producto sustitutivo o los datos de las primeras ventas de un producto similar. Por otro lado, también se pueden considerar a nivel cualitativo la mayoría de situaciones que se catalogan como poco claras; en el ejemplo mencionado, se puede cuantificar el cambio de temperatura que se espera y realizar en base a ello una predicción cuantitativa de las ventas.

También se recurre a métodos cualitativos cuando se desea interpretar y verificar si los resultados obtenidos al realizar previsiones con métodos cuantitativos son adecuados. Los métodos cuantitativos, presentados a continuación, pueden llegar a ser muy precisos; aun así, juzgar los resultados obtenidos con métodos cualitativos permite detectar anomalías y ambigüedades en los resultados.

Los métodos cuantitativos se basan en el análisis de los datos históricos y, en algunos casos, de los acontecimientos futuros conocidos para establecer un modelo de previsión. Dentro de los métodos cuantitativos se distinguen los de series temporales y los de causalidad o asociativos.

Los modelos de series de tiempo son modelos estadísticos que examinan datos pasados y predicen datos futuros en base a las tendencias obtenidas del análisis de datos históricos. Por otro lado, los modelos de causalidad determinan las asociaciones entre la variable a predecir y el resto de variables de las que se dispone en el histórico de datos y realizan las predicciones en base a estas asociaciones. Para ello, es útil la utilización de modelos de aprendizaje automático.

En la mayoría de los casos, los métodos cuantitativos proporcionan resultados más precisos y con menor varianza que los métodos cualitativos. Se debe considerar que la obtención de un modelo cuantitativo con una precisión elevada requiere tiempo de diseño y de actualización, recursos que puede que no sea oportuno invertir para todos los productos de los que se desea predecir las ventas. De todas formas, la obtención de un modelo con una

precisión aceptable puede requerir un tiempo parecido o menor que el necesario para obtener una previsión basada en métodos cualitativos. Por ello, en el caso de las previsiones de ventas, se considera que es oportuno la aplicación de métodos cuantitativos; incluso en casos considerados a priori poco claros o en los que no se dispone de un histórico de datos particular. Se valora, eso sí, la importancia de aplicar criterios cualitativos para juzgar los resultados proporcionados por los modelos cuantitativos.

En los apartados consecutivos se presentan modelos cuantitativos de previsión; tanto modelos estadísticos basados en series temporales como modelos de previsión de causalidad basados en el aprendizaje automático. Cabe remarcar que no existe un modelo único que sea en todos los casos el más adecuado para la previsión de ventas y, por ello, se considera oportuno el análisis de diversos modelos.

### **4.3. Modelos de previsión basados en el análisis de series temporales**

Como ya se ha mencionado, los modelos de previsión basados en análisis de series de tiempo son modelos estadísticos que predicen datos futuros en base a las tendencias de datos pasados. A continuación, se presentan ciertas características de las series de tiempo y diversos modelos de previsión basados en el análisis de éstas.

#### **4.3.1. Introducción a las series temporales**

Las series temporales son registros cronológicos de observaciones de una misma variable en intervalos de tiempo regulares. Un ejemplo de serie de tiempo son el registro de las unidades de venta semanales vendidas por un establecimiento.

Los factores que componen una serie de tiempo y que definen las variaciones de la variable objetivo o variable de interés son: tendencia secular, variación estacional, variación cíclica y variación irregular [6].

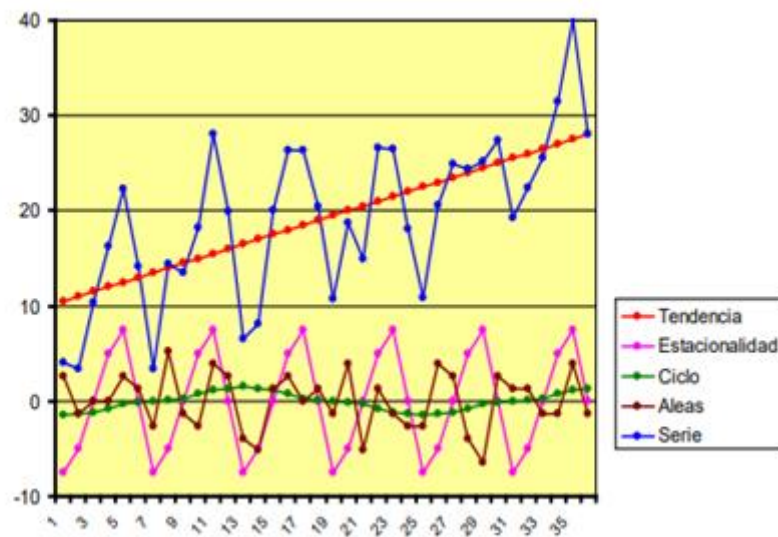


Figura 4.1 Componentes de las series de tiempo Fuente: [5]

La tendencia secular o tendencia a largo plazo es el componente principal de una serie temporal y es el resultado de los factores a largo plazo. Se caracteriza por un patrón regular que muestra la constancia, crecimiento o reducción de la variable de interés a lo largo de los diferentes períodos.

Las variaciones estacionales son las variaciones observadas a corto plazo que se deben a factores estacionales que ocurren año tras año en los mismos períodos; como, por ejemplo, las variaciones debidas al clima o ciertas festividades.

Las variaciones cíclicas son oscilaciones a largo plazo que se observan en las series de tiempo una vez eliminadas la tendencia y la estacionalidad. Un ejemplo de estas oscilaciones son los ciclos comerciales cuyas variaciones se deben a la prosperidad o depresión de la economía y no a factores estacionales como el clima o festividades.

Las variaciones irregulares se deben a cambios inesperados en la serie de tiempo que son poco probables de repetirse y no pueden ser explicados por la tendencia o las variaciones estacionales y cíclicas. Algunas de las causas de estas variaciones pueden ser huelgas, elecciones o terremotos.

Teniendo en cuenta los diferentes componentes de las series de tiempo, a nivel matemático estas se representan como [7]:

$$y_t = f(T_t, S_t, E_t) \quad (\text{Ec. 4.1})$$

donde:

- $y_t$  es el valor de la variable temporal en el período  $t$ .
- $T_t$  es el componente determinista de la tendencia-ciclo.
- $S_t$  es el componente determinista de la estacionalidad.
- $\epsilon_t$  es el componente irregular o residual.

La forma de la función  $f(\cdot)$  depende del método de descomposición utilizado. Los dos planteamientos más comunes son el aditivo y el multiplicativo. Ambos planteamientos consideran que la serie temporal es de tendencia estacionaria; es decir, que al eliminar los componentes deterministas la variable observada es estacionaria. Por un lado, el enfoque aditivo [8] asume que el valor de la variable temporal es suma del resto de componentes:

$$y_t = T_t + S_t + \epsilon_t \quad (\text{Ec. 4.2})$$

Por otro lado, el enfoque multiplicativo [8] considera que la variable observada es producto de los demás componentes:

$$y_t = T_t \cdot S_t \cdot \epsilon_t \quad (\text{Ec. 4.3})$$

Tal y como se observa en la Tabla 4.1, el planteamiento aditivo es apropiado cuando la magnitud de las fluctuaciones estacionales no varían al variar el nivel de la tendencia. De lo contrario, el enfoque multiplicativo es adecuado si las variaciones estacionales aumentan o disminuyen al crecer o decrecer la tendencia.

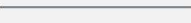
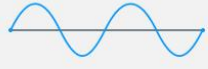
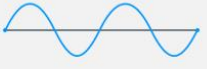



	NONSEASONAL	ADDITIVE SEASONAL	MULTIPLICATIVE SEASONAL
Constant Level	(Simple)  NN	 NA	 NM
Linear Trend	(HOLT)  LN	 LA	(WINTERS)  LM

Tabla 4.1 Representación de la estacionalidad aditiva y multiplicativa

Fuente: [8]

En los apartados consecutivos se presentan diversas variaciones de los modelos de previsión basados en el análisis de series temporales más utilizados: los modelos autorregresivos integrados de medias móviles (ARIMA) y los modelos de alisado exponencial. El modelo ARIMA y las variaciones de este pretenden identificar las autocorrelaciones existentes en los datos de las series de tiempo; mientras que los modelos de alisado exponencial se basan en describir la tendencia y la estacionalidad de las series analizadas.

### 4.3.2. Modelo autorregresivo

En un modelo autorregresivo, o *autorregressive model* (AR) en inglés, se predice la variable de interés como combinación lineal de los valores pasados de dicha variable [9]. El término autorregresivo indica que la variable se predice como regresión de sí misma. La expresión matemática de un modelo autorregresivo de orden  $p$ , modelo AR ( $p$ ), es:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (\text{Ec. 4.4})$$

### 4.3.3. Modelo de medias móviles

El modelo de medias móviles, o *moving average* (MA) en inglés, predice el valor de la variable objetivo como combinación lineal de predicciones pasadas [9]. A nivel matemático, un modelo de medias móviles de orden  $q$ , modelo MA ( $q$ ), se expresa como:

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (\text{Ec. 4.5})$$

Otras variaciones del modelo de medias móviles son el modelo de medias móviles simple (*Simple Moving Average* (SMA)) o el modelo exponencial de medias móviles (*Exponential Moving Average* (EMA)). Por ejemplo, el modelo SMA de orden  $n$  predice la variable de interés calculando la media aritmética de dicha variable en los últimos  $n$  períodos.

### 4.3.4. Modelo autorregresivo de medias móviles

Los modelos autorregresivos de medias móviles, o *autorregressive moving average models* (ARMA) en inglés, predicen la variable de interés de una serie temporal gracias a la suma de dos polinomios. El primero de ellos es un polinomio autorregresivo de orden  $p$  y el segundo un polinomio de medias móviles de orden  $q$  [9]. Por tanto, la expresión matemática de un modelo ARMA ( $p, q$ ) es:

$$y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (\text{Ec. 4.6})$$

Donde:

- $\phi_i$  son los parámetros del modelo autorregresivo y  $p$  el orden del polinomio de dicho modelo
- $\theta_i$  son los parámetros del modelo de medias móviles y  $q$  el orden del polinomio de dicho modelo
- $c$  es la constante del modelo
- $\epsilon_t$  son los términos del error



El modelo ARMA, al igual que los modelos autorregresivos y de medias móviles, no predice la variación de la tendencia ni de la estacionalidad; por tanto, es adecuado para predecir series de tiempo estacionarias. Una serie temporal estacionaria es aquella que su propiedad no depende del tiempo en que la serie es observada. Por ello, son series estacionarias aquellas que no tienen ni tendencia ni estacionalidad.

#### 4.3.5. Modelo autorregresivo integrado de medias móviles

El modelo autorregresivo integrado de medias móviles, o *autorregressive integrated moving average* (ARIMA) en inglés, puede ser adecuado para series temporales no estacionarias; en concreto, predice el componente de la tendencia, pero no las variaciones estacionales. Al igual que el modelo ARMA es una combinación del modelo autorregresivo y del de medias móviles; e incluye, además, tal y como indica el término “integrado”, el concepto de diferenciación de los datos [10].

La diferenciación de los datos consiste en calcular las diferencias entre observaciones consecutivas y es una manera de convertir series de tiempo no estacionarias en estacionarias. El modelo ARIMA permite diferenciar los datos de la serie de tiempo original más de una vez con el fin de conseguir una serie temporal estacionaria.

El modelo ARIMA se puede expresar como

$$y'_t = c + \epsilon_t + \sum_{i=1}^p \phi_i y'_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (\text{Ec. 4.7})$$

donde  $y'_t$  son los valores diferenciados de la serie temporal. Dicho modelo se puede nombrar como modelo ARIMA ( $p, d, q$ ), donde:

- $p$  es el orden de la parte autorregresiva
- $d$  es el orden de la diferenciación o el número de veces que los datos han sido diferenciados
- $q$  es el orden de la parte de medias móviles

Si el valor del orden  $d$  es igual a uno, los datos de la serie original se diferencian siguiendo la siguiente expresión matemática:

$$y'_t = y_t - y_{t-1} \quad (\text{Ec. 4.8})$$

Cabe remarcar que, por tanto, la serie temporal constará de  $T - 1$  valores, ya que no se puede calcular el valor diferenciado para la primera observación. Como ya se ha comentado, este modelo permite predecir correctamente series temporales con tendencia y sin estacionalidad; si la serie diferenciada presenta todavía una componente de tendencia al diferenciar los datos una vez, se debe aplicar un orden  $d$  superior. Por ejemplo, si se considera un orden  $d = 2$ , el

modelo ARIMA diferenciará los datos de la serie temporal original dos veces, aplicando la siguiente expresión matemática y obteniendo  $T - 2$  valores diferenciados:

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \quad (\text{Ec. 4.9})$$

#### 4.3.6. Modelo estacional autorregresivo integrado de medias móviles

El modelo estacional autorregresivo integrado de medias móviles, o *autorregressive integrated moving average* (SARIMA) en inglés, incluye además de los parámetros del modelo ARIMA una serie de términos que permiten predecir la variación estacional de una serie temporal [11]. Por tanto, el modelo SARIMA permite predecir de manera adecuada series de tiempo con tendencia y estacionalidad.

El modelo SARIMA  $(p, d, q) \times (P, D, Q)_m$  incluye los órdenes  $(p, d, q)$  que permiten predecir, al igual que en el modelo ARIMA, el componente determinista tendencia-ciclo de la serie temporal y los órdenes  $(P, D, Q)_m$  que predicen el componente determinista de la estacionalidad [12].  $(p, d, q)$  son respectivamente el orden de la parte autorregresiva, de la diferenciación y de la parte de medias móviles utilizados para predecir la componente tendencia-ciclo; y  $(P, D, Q)$  corresponden a los órdenes de la parte autorregresiva, de la diferenciación y de la parte de medias móviles que predicen la estacionalidad. El valor  $m$  son el número de observaciones anuales que se dispone de la serie temporal. Por ejemplo, si se poseen datos semanales el valor  $m$  será 52.

En la Figura 4.1 se muestra la expresión matemática del modelo SARIMA  $(1,1,1) \times (1,1,1)_4$  considerando estacionalidad multiplicativa.

$$(1 - \phi_1 B) (1 - \Phi_1 B^4) (1 - B) (1 - B^4) y_t = (1 + \theta_1 B) (1 + \Theta_1 B^4) e_t.$$

↑

(Non-seasonal)  
AR(1)

↑

(Non-seasonal)  
difference

↑

(Non-seasonal)  
MA(1)

↑

(Seasonal)  
AR(1)

↑

(Seasonal)  
difference

↑

(Seasonal)  
MA(1)

Figura 4.2 Componentes del modelo SARIMA(1,1,1)x(1,1,1)<sub>4</sub> con estacionalidad multiplicativa

Fuente: [13]

#### 4.3.7. Modelo estacional autorregresivo integrado de medias móviles con entrada exógena

El modelo estacional autorregresivo integrado de medias móviles con entrada exógena, o *autorregressive integrated moving average with exogenous input* (SARIMAX) en inglés, es una variación del modelo SARIMA el cual incluye una regresión que pretende determinar la relación entre la variable de interés con respecto a otras variables exógenas (X) [14].

#### 4.3.8. Modelo de alisado exponencial simple

El modelo de alisado exponencial simple predice la variable de interés de una serie de tiempo a través de medias ponderadas donde el peso de las ponderaciones decrece exponencialmente con la antigüedad de los datos [15]. Este modelo es adecuado para aquellas series temporales sin un claro patrón de tendencia o estacionalidad. La expresión matemática del modelo para el cálculo de la variable de interés consecutiva ( $T + 1$ ) es:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots \quad (\text{Ec. 4.10})$$

Donde:

- $\hat{y}$  es la predicción de la variable objetivo de la serie temporal
- $y$  son los valores registrados de las series temporales
- $T$  es el número de registros de la variable de interés
- $\alpha$  es un valor entre [0,1] y controla el decrecimiento de las ponderaciones

#### 4.3.9. Modelo de alisado exponencial doble

El modelo del alisado exponencial doble permite predecir series temporales con tendencia. El modelo matemático está formado por una ecuación de predicción y dos ecuaciones de alisado exponencial [15]. Por tanto, la expresión para calcular la variable ( $T + h$ ) es:

$$\text{Ecuación de predicción} \quad \hat{y}_{T+h|T} = l_t + h \cdot b_t \quad (\text{Ec. 4.11})$$

$$\text{Alisado exponencial nivel} \quad l_t = \alpha y_T + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (\text{Ec. 4.12})$$

$$\text{Alisado exponencial tendencia} \quad b_t = \beta(l_t + l_{t-1}) + (1 - \beta)b_{t-1} \quad (\text{Ec. 4.13})$$

Donde  $\alpha$  y  $\beta$  son los parámetros de las ecuaciones de alisado exponencial y son valores entre [0,1]. Existe también el modelo de alisado exponencial triple que permite predecir series temporales con estacionalidad, ya sea multiplicativa o aditiva, y el que incluye un tercer alisado exponencial para predecir la componente de la estacionalidad.

## 4.4. Modelos de previsión de causalidad basados en el aprendizaje automático

Los modelos de previsión de causalidad determinan relaciones causa-efecto entre los datos históricos y las proyectan para obtener previsiones. A continuación, se presentan diversos modelos que utilizan el aprendizaje automático para detectar dichas relaciones.

### 4.4.1. Introducción al aprendizaje automático

El aprendizaje automático, o *machine learning* (ML) en inglés, es una aplicación de la inteligencia artificial que proporciona a sistemas la habilidad de automáticamente aprender y mejorar en base a la experiencia sin ser explícitamente programados [16].

Los modelos de *machine learning* se basan en algoritmos de aprendizaje automático que se entrenan en base a datos históricos para que el modelo responda a las necesidades de una casuística concreta. Los algoritmos de ML se describen como una función objetivo de aprendizaje  $f$  que pretende, en base a unas entradas  $X$ , explicar una variable de interés  $y$  [17]:

$$y = f(X) \quad (\text{Ec. 4.14})$$

El aprendizaje automático es utilizado para definir la función  $f$  cuando esta es desconocida. Datos históricos son usados para entrenar los algoritmos de ML y conocer una función  $f$  que permita explicar la variable de interés  $y$  en base a las entradas  $X$ .

En los modelos de previsión, el aprendizaje automático es utilizado para realizar predicciones de la variable de interés en base a entradas  $X$  futuras. Los algoritmos de ML son diferentes técnicas que permiten estimar la función objetivo  $f$ . Diferentes algoritmos realizan diferentes hipótesis sobre la forma y la estructura de la función objetivo, y sobre cómo optimizarla para obtener una función  $f$  lo más adecuada posible. Debido a los diferentes enfoques que proponen la gran diversidad de algoritmos de ML, es importante que para cada casuística concreta se entrenen diferentes algoritmos con el fin de construir un modelo definitivo en base a aquellos algoritmos que sean los más adecuados en cada caso.

Con el fin de entrenar los algoritmos de aprendizaje automático y conocer cómo de adecuados son para explicar una casuística concreta, existen diferentes planteamientos. Uno de los más comunes es utilizar un 70-80% de los datos históricos de que se dispone para entrenar los algoritmos y el 20-30% de los datos restantes para valorar con que precisión permite el modelo

generado adaptarse a nuevas entradas. Esta división permite valorar si el modelo obtenido explica de manera adecuada los datos o, de si lo contrario, el modelo presenta sobreajuste o no se ajusta a los datos.

Con el fin de obtener un modelo que pueda predecir correctamente la variable de interés dadas nuevas entradas, es importante que el modelo sea robusto (*robust* en inglés). Si de lo contrario el modelo explica con demasiada precisión (*overfitting* en inglés) los datos históricos con los que se ha entrenado [18], este no explicará adecuadamente datos futuros. En la Figura 4.3 se representan estos conceptos.

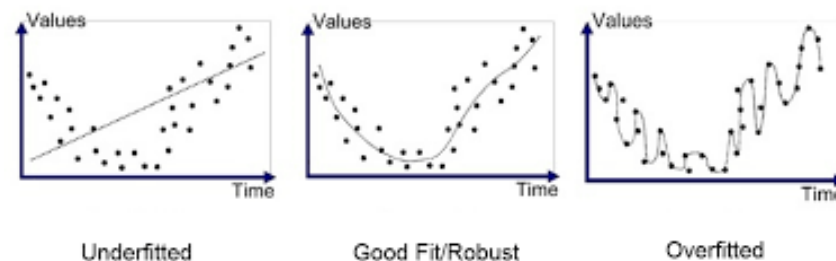


Figura 4.3 Representación de los conceptos de underfitted, good fit y overfitted Fuente: [18]

Los modelos de previsión de causalidad basados en el aprendizaje automático consideran algoritmos de ML entrenados con datos históricos y permiten detectar relaciones de causa-efecto en dichos datos y proyectarlas en el futuro obteniendo así las previsiones. A continuación, se presentan diferentes algoritmos, agrupados por similitud, que pueden ser utilizados para realizar previsiones de causalidad.

#### 4.4.2. Algoritmos de regresión

Los algoritmos de regresión o *regression algorithms* pretenden modelar la relación entre variables y refinan el modelo gracias al cálculo iterativo del error en las predicciones realizadas sobre los datos de entrenamiento.

El objetivo de estos algoritmos de aprendizaje automático es obtener un modelo que permita realizar predicciones sobre datos desconocidos, que no forman parte de los datos de entrenamiento. Esto diferencia a los modelos de ML basados en los algoritmos de regresión del concepto estadístico de regresión donde se busca encontrar la relación entre diferentes variables respecto a una variable de interés conocidos los datos [19].

Algunos de los algoritmos de regresión son:

- Regresión lineal o *Linear Regression*: predice la variable de interés  $y$  como combinación lineal de las variables  $X$  [20]. Por tanto, supone que la función  $f$  es de la

forma:

$$y = \theta_0 + \theta_1 \cdot X \quad (\text{Ec. 4.15})$$

Uno de las maneras de predecir los coeficientes  $\theta_0$  y  $\theta_1$  es minimizar la función de coste  $J$  que penaliza los errores en las predicciones realizadas por el modelo sobre los datos de entrenamiento utilizando el algoritmo de descenso de gradiente o *gradient descent*. Donde la función coste se expresa como:

$$J = \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2 \quad (\text{Ec. 4.16})$$

El algoritmo del descenso de gradiente permite encontrar un mínimo local de dicha función iniciando los valores de los coeficientes con un valor aleatorio y actualizándolos hasta que la función converja según:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad \text{para } j = 0 \text{ y } j = 1 \quad (\text{Ec. 4.17})$$

Donde el valor del coeficiente  $\alpha$  debe ser escogido para que la convergencia no sea muy lenta (valores  $\alpha$  demasiado bajos) ni impida que no converja (valores  $\alpha$  demasiado altos).

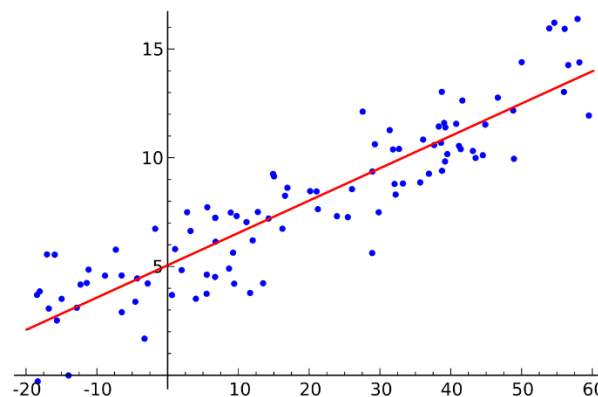


Figura 4.4 Regresión lineal Fuente: [20]

- Curvas de regresión adaptativa multivariante o *Multivariate Adaptive Regression Splines* (MARS): es una extensión del algoritmo de regresión lineal que incorpora no linealidades e interacciones entre variables [21]. Puede expresarse como:

$$y = \theta_0 + \sum_{i=1}^k \theta_i B(X_i) \quad (\text{Ec. 4.18})$$

Donde  $\theta_i$  son los parámetros para las diferentes variables consideradas ( $X_i$ ) y  $B(X_i)$  puede ser, por ejemplo, una constante o de la forma  $\max(0, X_i - \text{constante})$ . Se

utilizan los algoritmos *forward/backward stepwise* para determinar los coeficientes  $\theta_j$  más adecuados.

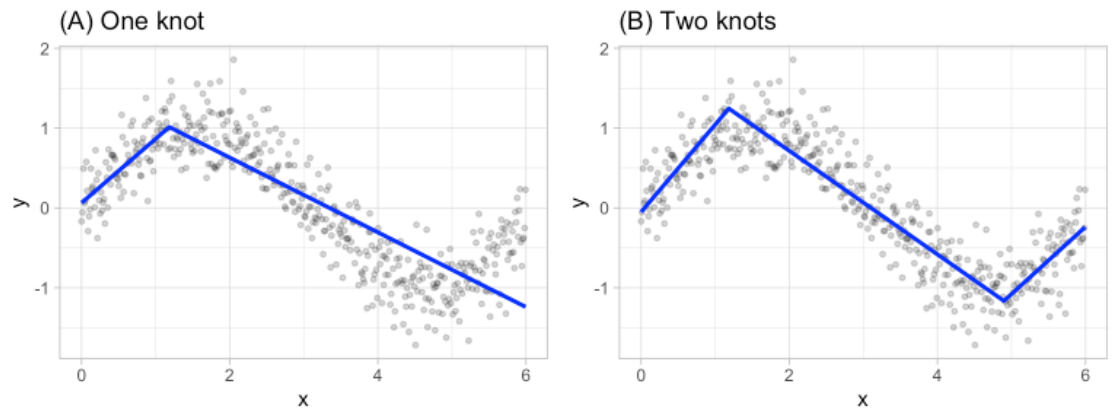


Figura 4.5 Splines de regresión adaptativa multivariante Fuente: [21]

- *Locally Estimate Scatterplot Smoothing* (LOESS): este algoritmo permite encontrar una curva para realizar predicciones dadas nuevas entradas [22]. El algoritmo ajusta los datos de entrenamiento a una curva a través de regresiones locales ponderadas.

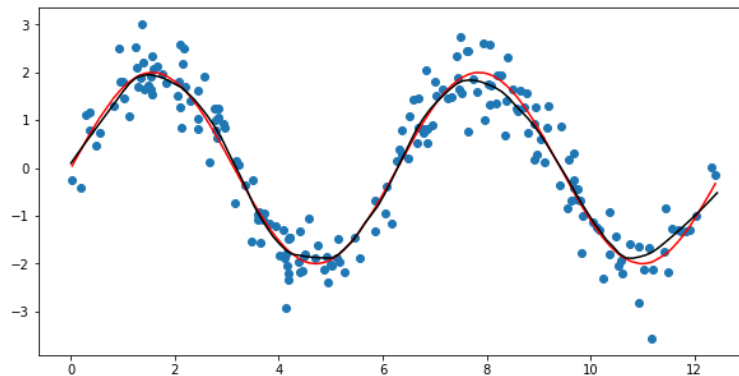


Figura 4.6 Locally Estimate Scatterplot Smoothing Fuente: [22]

#### 4.4.3. Algoritmos basados en casos

Los algoritmos basados en casos o *instance-based algorithms* en inglés comparan nuevos casos con casos vistos en los datos de entrenamiento y que han sido almacenados en la memoria. Debido a que no generan conocimiento explícito, estos algoritmos son denominados también como de aprendizaje perezoso (*lazy learning*) [23].

Uno de los algoritmos más comunes de esta tipología, y que puede utilizarse para realizar previsiones, es el k-vecinos más cercanos o *k-Nearest Neighbor* (kNN). Este algoritmo almacena los valores de la variable de interés y de todas las variables  $X$  de todos los casos

de los datos de entrenamiento. A la hora de decidir el valor de la variable  $y$  dada una nueva entrada  $X$ , el algoritmo realiza la media de los valores de la variable de interés de aquellos  $k$  casos que tienen unos valores  $X$  más parecidos, considerando una medida de similitud, a la nueva entrada.

En el ejemplo de la Figura 4.7 se calcula la previsión del valor doceavo considerando el algoritmo de los tres vecinos más cercanos. El valor de la variable de interés 12 se calcula como media de los tres valores conocidos de los casos más similares; se calcula como media del valor  $y$  de los datos registrados como 1, 5 y 6.

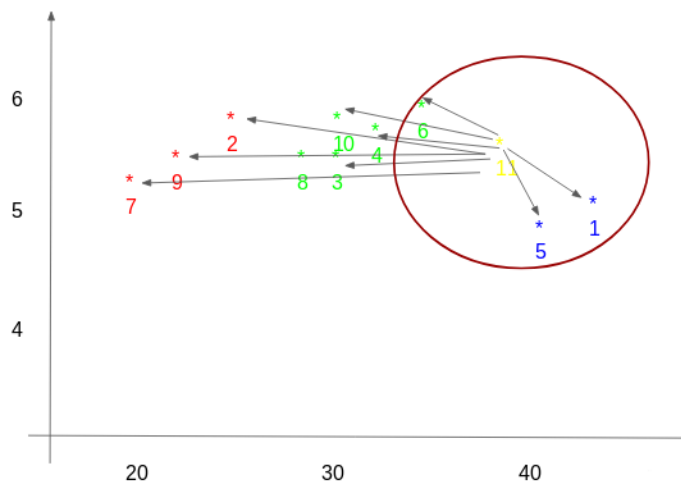


Figura 4.7 Ejemplo de aplicación de  $k$ -vecinos más cercanos con  $k=3$  Fuente: [24]

Las medidas de similitud más utilizadas se basan en funciones de distancia como son la distancia euclidiana o la distancia de Manhattan [25]. Considerando

$$x_i = (x_{1i}, x_{2i}, \dots, x_{si}) \in X, \quad (\text{Ec. 4.19})$$

dichas distancias entre dos puntos  $x_i$  y  $x_j$  pueden expresarse como:

$$\sqrt{\sum_{r=1}^s (x_{ri} - x_{rj})^2} \quad (\text{Ec. Distancia euclidiana}) \quad (\text{Ec. 4.20})$$

$$\left( \sum_{r=1}^s (|x_{ri} - x_{rj}|)^p \right)^{1/p} \quad (\text{Ec. Distancia de Minkowski}) \quad (\text{Ec. 4.21})$$

#### 4.4.4. Algoritmos basados en árboles de decisión

Los algoritmos basados en árboles de decisión también pueden ser utilizados para realizar



previsiones. Estos algoritmos crean modelos basándose en datos de entrenamiento y aprendiendo de ellos reglas de decisión simples con estructura de árbol que permiten predecir el valor de la variable de interés dadas nuevas entradas. Estos algoritmos son generalmente rápidos y precisos.

Las decisiones se toman siguiendo una estructura de árbol que parte de un nodo raíz (*root node*) y considera diferentes nodos de decisión (*decision node*) hasta que un nodo terminal (*terminal node*) estima el valor de la variable de interés [26].

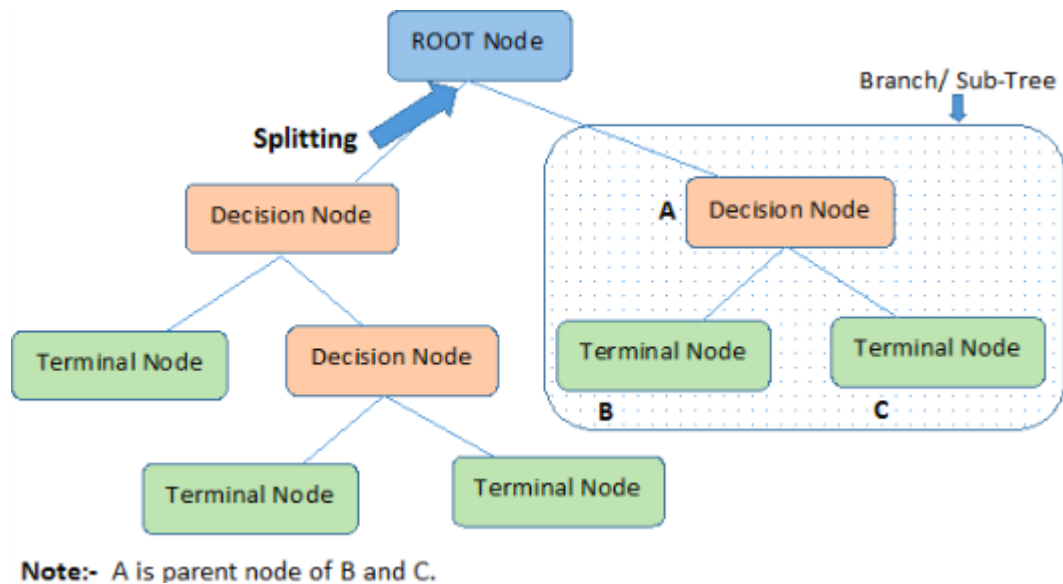


Figura 4.8 Estructura de los algoritmos basados en árboles de decisión Fuente: [27]

Los algoritmos basados en árboles de decisión utilizados para realizar previsiones de variables continuas, discretizan la variable de interés previamente a construir el modelo. Estos algoritmos utilizan otros algoritmos para decidir si convertir un nodo en un nodo de decisión o en un nodo terminal. Algunos de los algoritmos que se utilizan son: ID3, C4.5, *Classification And Regression Tree* (CART) o *Multivariate Adaptive Regression Splines* (MARS).

En concreto, el algoritmo ID3 realiza una búsqueda *greedy*, escoge la opción óptima en cada punto de decisión, de arriba hacia abajo sin considerar la posibilidad de retroceso. Es decir, en el nodo raíz parte considerando la totalidad de los datos de entrenamiento y en cada iteración el algoritmo considera los atributos no explorados de  $X$  y calcula la entropía (entropy ( $H$ )) y la ganancia de información (information gain ( $IG$ )) para dichos atributos. A continuación, selecciona el atributo con menor valor de  $H$  o con mayor valor  $IG$  y divide los datos de entrenamiento en función del atributo en subconjuntos de datos. El algoritmo recorre cada subconjunto de datos realizando el mismo procedimiento y seleccionando atributos todavía no considerados. La entropía se expresa como:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (\text{Ec. 4.22})$$

Donde  $S$  es el conjunto o subconjunto de datos de entrenamiento considerados y  $p_i$  es la probabilidad de que el atributo considerado tome el valor indexado como  $i$ . Por otro lado, la ganancia de información puede expresarse como:

$$\text{Information Gain} = \text{Entropy}(\text{before}) - \sum_{j=1}^K \text{Entropy}(j) \quad (\text{Ec. 4.23})$$

Donde *before* es el conjunto de datos considerados antes de la división de los datos por el atributo analizado y  $K$  es el número de subdivisiones de datos que se realizarán.

#### 4.4.5. Algoritmos basados en redes neuronales

Los algoritmos basados en redes neuronales pretenden imitar el comportamiento de los axones de las neuronas del cerebro biológico y pueden ser también utilizados para crear modelos de previsión.

Estos algoritmos consideran un sistema de conexiones donde cada unidad simple o neurona artificial está conectada a muchas otras. Las neuronas se agrupan en capas y, generalmente, las redes neuronales se basan en una capa de entradas, con tantas neuronas como variables de entrada  $X$ ; un número determinado de capas intermedias o capas ocultas, la dimensión de la cual se fija durante el diseño de la red; y una última capa de salida que contiene tantas neuronas como variables de salida  $y$  [28].

El funcionamiento de la red neuronal se basa en que cada neurona de la primera capa está conectada con todas las neuronas de la segunda capa. De igual manera, cada neurona de la segunda capa está conectada con las neuronas de la siguiente capa, ya sea una capa intermedia o la capa de salida; y así hasta llegar a la capa de salida. Cada enlace entre neuronas de las distintas capas tiene asociado un peso que indica la importancia de la conexión y que multiplica el valor de la neurona anterior y se suma con el resto de conexiones para obtener la entrada que se aplicará a la función de activación para obtener el valor final de la neurona a la que conecta. De esta manera y decidiendo, gracias a los datos de entrenamiento, los pesos adecuados para cada una de las conexiones, se puede predecir la variable de salida  $y$  a partir de nuevas variables de entrada  $X$ .

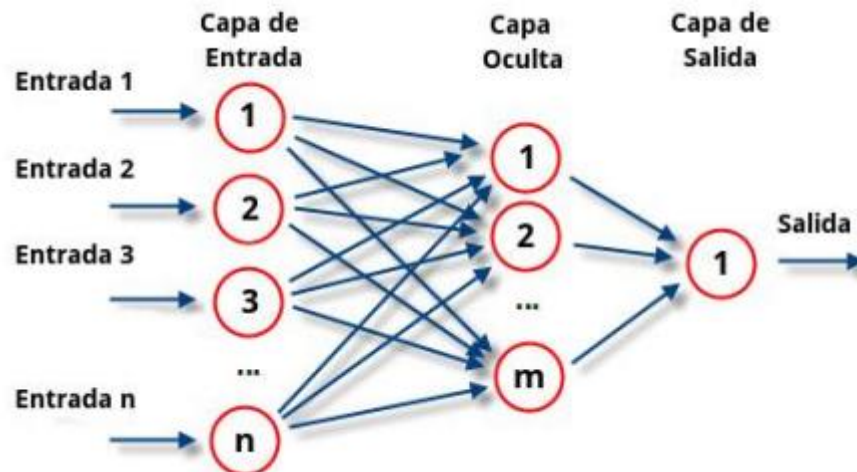


Figura 4.9 Esquema de una red neuronal de una capa oculta

Fuente: commons.wikipedia.org

Existen diferentes funciones de activación que pueden ser utilizadas, algunas de las más populares son:

$$\text{Rectified Linear Units (ReLU)} \quad \max(0, z) \quad (\text{Ec. 4.24})$$

$$\tanh \quad \tanh(z) \quad (\text{Ec. 4.25})$$

$$\text{Sigmoid activation} \quad 1/(1 + e^{-1 \cdot z}) \quad (\text{Ec. 4.26})$$

#### 4.4.6. Algoritmos de conjunto

Los modelos basados en algoritmos de conjunto están compuestos por múltiples modelos que son entrenados independientemente y que sus predicciones son combinadas para formar una predicción conjunta, con el fin de reducir la varianza y aumentar la precisión de los resultados [29]. Estos modelos también pueden ser utilizados para realizar predicciones, dos de los más utilizados son:

- **Bosques aleatorios o *Random Forest*:** el modelo basado en este algoritmo considera diversos árboles y el valor de las predicciones se calcula como el promedio de las salidas de los diferentes árboles.

En concreto, se generan un número determinado de árboles  $T$  considerando, cada uno de ellos, un subconjunto aleatorio de  $N$  casos de los datos de entrenamiento y un subconjunto de variables predictoras  $m$  por nodo. La variable predictora que proporciona la mejor división en cada nodo, de acuerdo con una función objetivo, es la utilizada para realizar la división del nodo. El nodo consecutivo realiza lo mismo considerando un subconjunto de variables predictoras distinto.

- *XGBoost (Extra gradient boosting)*: este modelo también se basa en diversos árboles de decisión; pero a diferencia del *Random Forest*, se potencian los resultados de los diferentes árboles, ya que los datos se procesan secuencialmente minimizando, en cada iteración, una función coste previamente definida.

## 5. Metodología para la selección y aplicación de un modelo de previsión de ventas

Como se ha visto en el apartado anterior, existen numerosos modelos de previsión y el modelo más adecuado varía para cada caso particular. Debido a ello, se considera oportuno presentar una metodología que permita obtener un modelo de previsión de ventas adecuado dado un caso concreto.

### 5.1. Precisión necesaria del modelo de previsión dependiendo de la tipología de producto

Tal y como se ha evaluado en el apartado de *La importancia de la previsión de ventas en la gestión de la cadena de suministro*, no todas las ventas de los diferentes productos requieren ser previstas con la misma precisión. En términos generales, los productos de alta rotación necesitan ser previstos con mayor precisión que los productos de baja rotación. Además, otros factores como el coste de almacenar el producto, el tiempo de abastecimiento y el precio del producto, deben ser considerados a la hora de establecer la previsión necesaria de cada artículo.

Por ello, se considera oportuno realizar un análisis de agregación o *cluster analysis*, para decidir qué productos se consideran críticos y necesitan una precisión de su previsión de ventas elevada. Dicho análisis debe contemplar los factores nombrados y considerar con mayor criticidad aquellos productos que generen un mayor número de ventas y/o ingresos, cuyo abastecimiento no sea reactivo y cuyo coste de almacenaje sea elevado.

Se recomienda seleccionar como productos críticos a aproximadamente un 20% del total de la cartera de productos; los cuales, siguiendo la regla de Pareto, serán responsables del 80% de beneficios y de una posible pérdida parcial de ellos debido a la baja precisión de sus modelos de previsión de ventas.

Para los productos considerados críticos, se recomienda definir un modelo de previsión tan preciso como sea posible, invirtiendo en él tiempo de diseño y de actualización. Por otro lado, para el resto de productos se considera que puede utilizarse uno menos robusto. Mientras que se considera necesario una precisión superior al 90% para los productos críticos, una precisión del 75-80% será suficiente para el resto de productos; dichas precisiones podrían variar dependiendo de cada casuística. Para todos los casos se considera oportuno, tal y como se ha mencionado a la hora de analizar los diferentes modelos de previsión existentes, aplicar métodos de previsión cuantitativos y valorar críticamente sus resultados mediante métodos cualitativos.

Tal y como también se ha expuesto, es oportuno considerar los mismos métodos de previsión para los productos nuevos que para el resto de productos. Debido a que no se dispondrá de un histórico de ventas para un producto nuevo, se debe emplear para los modelos de previsión históricos de un producto sustituto o generalizar los comportamientos de nuevos productos gracias a datos pasados.

Cabe destacar que, en caso de disponer de una cartera de productos con un número elevado de artículos diferentes, será oportuno considerar, a la hora de hacer las previsiones de las ventas, familias de productos en vez de productos únicos. Una familia es el conjunto de productos que responden a una misma necesidad genérica del consumidor.

Por tanto, se debe aplicar la metodología presentada en los apartados consecutivos para realizar las previsiones de venta de toda la cartera de productos. De todas maneras, con el fin de disminuir el tiempo de diseño, se aconseja considerar por separado los productos críticos de los no críticos e invertir mayores recursos en la etapa de diseño de los primeros con el fin de conseguir una precisión superior en estos. La metodología que se presenta considera invertir tiempo en definir los parámetros del modelo que permiten mejores resultados y en considerar diferentes modelos candidatos con el fin de obtener una mejor precisión en el modelo definitivo. Estas etapas son imprescindibles para definir modelos de productos críticos y disponer de una precisión superior al 90%. Para estos productos, se considerará también la posibilidad de disponer de diversos modelos de previsión para las distintas familias con el fin de aumentar la precisión. Por otro lado, para obtener la precisión deseada para los productos no críticos puede ser suficiente disponer de unos parámetros aceptables, considerar un solo modelo candidato y utilizar un único modelo de previsión definitivo para la totalidad de los productos.

## **5.2. Exploración y manipulación de los datos**

Para crear un modelo de previsión es necesario comenzar explorando los datos y entender en el mayor detalle posible la casuística concreta a la que se quiere dar solución. A continuación, se considera oportuno manipular los datos con el fin de utilizar aquellos que permitirán obtener una solución concisa y en un tiempo de cálculo que se considere aceptable.

Dependiendo de la casuística concreta que se desee analizar, la exploración y manipulación de datos puede variar. De todas formas, en cualquier caso, será adecuado realizar una representación gráfica de los datos, detectar y eliminar datos anómalos y realizar una selección de las variables explicativas, a parte de la variable de interés, que se desean

analizar.

Con el fin de seleccionar las variables adecuadas que deberían considerarse en el modelo de previsión se puede realizar un análisis de las correlaciones entre las variables existentes en los datos disponibles y no seleccionar aquellas variables correlacionadas entre sí para formar parte del modelo. También, puede ser adecuado proponer la recogida de nuevos datos para analizar nuevas variables o considerar variables que sean combinación de otras; cuanto más se explore y se analice la casuística concreta, más correcta será la selección de variables.

### **5.3. División de los datos para la valoración de la precisión del modelo**

Debido a que la gran mayoría de modelos cuantitativos necesitan de unos datos de entrenamiento para definirse, se debe destacar la importancia de dividir los datos de los que se dispone para por un lado entrenar el modelo y por el otro lado conocer la precisión de este. Existen diversos enfoques y maneras de dividir los datos, siendo la división de los datos en datos de entrenamiento y prueba una de las posibilidades.

#### **5.3.1. División de los datos en datos de entrenamiento y de prueba**

Una de las opciones es dividir el histórico de datos del que se dispone en dos: utilizando un 70-80% como datos de entrenamiento y un 20-30% como datos de prueba.

Los datos de entrenamiento (*training data*) son aquellos datos utilizados para entrenar los algoritmos y definir el modelo. Por otro lado, los datos de prueba (*test data*) se utilizan para valorar cómo se comporta el modelo dadas nuevas entradas y conocer así cual es la precisión que puede esperarse al utilizarse dicho modelo para realizar previsiones.

Además, la división de los datos en datos de entrenamiento y de prueba, también permite conocer si el modelo presenta sobreajuste respecto los datos de entrenamiento (concepto de *overfitting* visto en la sección 4.4.1). El modelo presentará sobreajuste si se observa que la precisión del modelo es mucho menor al realizar previsiones sobre los datos de prueba que sobre los datos de entrenamiento. Si se detecta *overfitting* existen técnicas para evitar dicho sobreajuste, como son reducir el número de variables explicativas o realizar el proceso de regularización (*regularization*), explicación del cual queda fuera del abasto del presente proyecto.

## 5.4. Aplicación de modelos candidatos y valoración de los resultados

Tal y como se ha comentado anteriormente, el modelo de previsión más adecuado dado un caso particular no siempre se basará en los mismos algoritmos y, por tanto, la exploración y consideración de diversos modelos es necesaria. Por ello, se propone analizar lo correctas que son las previsiones dados diversos modelos con el fin de valorar los resultados y proponer, en base a ellos, un modelo definitivo adecuado. Por tanto, tal y como se propone a continuación, se diseñarán diversos modelos candidatos.

### 5.4.1. Selección de un modelo candidato

Para seleccionar un modelo candidato se recomienda basarse en la literatura y en la propia experiencia con el fin de analizar aquellos modelos que potencialmente serán más adecuados para el caso al que se desea dar respuesta. En el presente proyecto, los modelos de previsión analizados son aquellos que potencialmente son adecuados para las previsiones de ventas.

### 5.4.2. Selección de los mejores parámetros para el modelo candidato

La gran mayoría de modelos se basan en algoritmos que consideran parámetros que pueden tomar diversos valores, y la correcta selección del valor de dichos parámetros determinará la precisión y adecuación del modelo. Por ejemplo, el modelo ARMA presentado en la sección 4.3.4 considera los parámetros  $p$  y  $q$  y el valor adecuado de estos parámetros dependerá de la casuística concreta que sea analizada.

Existen técnicas que permiten examinar y determinar valores adecuados de los parámetros dados unos datos de entrenamiento. A continuación, se presentan las técnicas más utilizadas: la búsqueda aleatoria y la búsqueda de cuadrículas [30]. Ambas técnicas generan el modelo en base a diferentes combinaciones de los valores de los parámetros, valores determinados en la fase de diseño, y establecen que combinación de valores permite unos mejores resultados. En concreto,

- la búsqueda aleatoria o *Random Search* considera combinaciones aleatorias de los valores de los parámetros. No se consideran todas las combinaciones de valores posibles, sino un número determinado de combinaciones aleatorias. Número establecido por el autor durante la etapa de diseño.
- Por otro lado, la búsqueda de cuadrículas o *Grid Search* considera todas las combinaciones posibles de los valores de los parámetros.



### 5.4.3. Entrenamiento y valoración del modelo candidato

Con el fin de establecer un modelo candidato, los algoritmos que lo conforman deben ser entrenados con los datos de entrenamiento que se dispone. A continuación, el modelo entrenado será considerado como el modelo candidato definitivo y será valorado evaluando la precisión de las previsiones dados las entradas de los datos de prueba.

Con el fin de valorar los distintos modelos candidatos, es importante que para todos ellos se utilice el mismo criterio de evaluación de la precisión de las previsiones; la precisión en la que se predice la variable de interés considerando las entradas  $X$  de los datos de prueba.

Una de las maneras más comunes para evaluar la precisión de las previsiones es calculando el error cuadrático medio o el error relativo de las previsiones de la variable de interés dadas las entradas de los datos de prueba respecto el valor real de la variable  $y$ .

## 5.5. Selección del modelo definitivo

Una vez evaluados diferentes modelos, se puede proceder a seleccionar el modelo definitivo. Para ello, existen dos enfoques, seleccionar como modelo definitivo aquel modelo candidato que haya realizado las previsiones de los datos de prueba con mayor precisión, o se puede crear un modelo definitivo que sea combinación de dos o más modelos candidatos. Este último concepto se presenta a continuación.

### 5.5.1. Métodos combinados de aprendizaje

Los métodos combinados de aprendizaje o *ensemble learning* combinan las decisiones, de en este caso previsiones, de diversos modelos para aumentar la precisión y disminuir la varianza del modelo definitivo [29]. Existen diferentes algoritmos que permiten crear un modelo definitivo combinando diversos modelos candidatos. Algunas de estas técnicas son:

- De media o *averaging*: las previsiones realizadas por el modelo definitivo son una media de las predicciones realizadas por los modelos candidatos considerados o por aquellos que presentan un mayor rendimiento, un menor error en las predicciones.
- De media ponderada o *weighted average*: las previsiones del modelo definitivo serán una media ponderada de dos o más de los modelos candidatos. Las ponderaciones dependerán de la importancia de cada uno de los modelos considerados para predecir la variable de interés.

## 5.6. Aplicación y actualización del modelo definitivo

Una vez definido el modelo definitivo este puede ser implementado y utilizado para realizar previsiones de la variable de interés y dadas nuevas entradas  $X$ . A la hora de utilizar las previsiones de ventas generadas, es recomendable revisar la adecuación de los resultados mediante métodos cualitativos. Por la naturaleza de estos métodos, dichas revisiones deben ser realizadas por expertos.

Se debe destacar también que todo modelo debe ser revisado y actualizado con regularidad a lo largo del tiempo. La frecuencia adecuada de dichas actualizaciones dependerá de cada casuística. Para actualizar un modelo puede ser adecuado seguir esta misma metodología presentada o realizar actualizaciones entrenando los algoritmos del modelo definitivo con datos de entrenamiento más recientes.

## 6. Aplicación y valoración de los métodos de ventas analizados en un caso práctico

A continuación, se presenta un caso práctico y se aplican y valoran los diversos métodos de ventas analizados aplicando la metodología presentada.

### 6.1. Presentación del caso práctico

Con el fin de aplicar los conocimientos y análisis expuestos en los apartados previos, en esta sección se pretende dar solución a un caso práctico. En concreto, se desea predecir las ventas de un establecimiento de la compañía Walmart, corporación multinacional de tiendas que operara grandes almacenes de venta al consumidor.

Analizando la cadena de suministro asociada al caso práctico, Walmart es el detallista que gracias a los diversos establecimientos (puntos de venta) que dispone oferta productos finales a sus clientes. Por este motivo, una adecuada predicción de las ventas en sus establecimientos permitirá realizar una gestión eficiente de las cadenas de suministro asociadas a los diferentes productos ofrecidos.

El establecimiento Walmart considerado en el caso práctico consta de 98 departamentos y el objetivo es desarrollar un modelo de previsión que permita estimar lo mejor posible las ventas de cada uno de ellos. Para ello se dispone del número de unidades vendidas en cada uno de los departamentos durante 143 semanas consecutivas, lo que equivale a 2 años y 9 meses. En la Figura 6.1 se muestran las ventas semanales para uno de los departamentos y para todas las semanas de las que se dispone de datos.

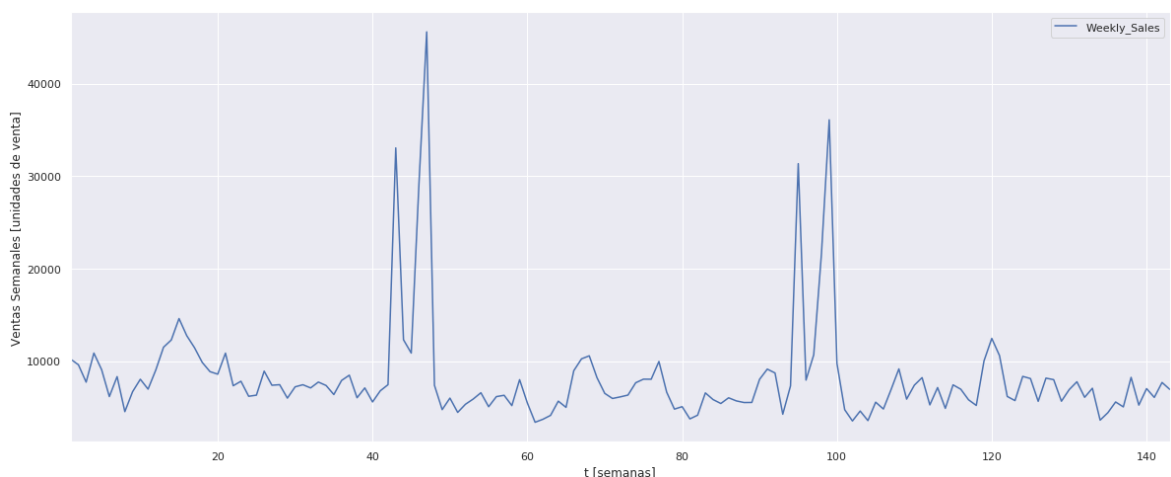


Figura 6.1 Ventas semanales de una de los departamentos del establecimiento Walmart

Fuente: Propia

Además, para cada una de las semanas de las que se dispone de las unidades vendidas, se conoce también:

- La fecha concreta del viernes de la semana considerada. Las fechas del histórico de datos van desde el 5 de febrero de 2010 hasta el 26 de julio de 2013.
- Si la semana es festiva o no.
- La temperatura ambiente media durante la semana.
- Si se ha realizado o no una acción comercial durante la semana y qué tipo de acción ha sido realizada. Se consideran 4 tipos de acciones comerciales diferentes; no se conocen los detalles de en qué consisten dichas acciones.
- El Índice de Precios al Consumidor (IPC), *Consumer Price Index* (CPI) en inglés, de la semana. Este índice mide la evolución del coste promedio de una cesta de bienes y servicios considerados representativos del consumo final de los hogares y se utiliza para evaluar el cambio de los precios asociados al coste de vida.
- La tasa de desempleo durante la semana considerada; dicho valor varía trimestralmente.

Como ya se ha mencionado, el caso práctico consiste en desarrollar un modelo que permita predecir las ventas para cada uno de los departamentos del establecimiento Walmart considerado. Este modelo estará basado en el histórico de datos del que se dispone y permitirá predecir las ventas futuras. Como es lógico, se desea que el modelo prevea las ventas lo mejor posible para aumentar así la eficiencia de la cadena de suministro asociada al establecimiento. En apartados posteriores se definirá cómo se evaluará la calidad del modelo a definir.

Cabe destacar que el modelo encontrado permitirá predecir las ventas a corto, medio y largo plazo. De todas maneras, se debe tener presente que dicho modelo debería ser actualizado a lo largo del tiempo con el fin de que su precisión no disminuya. En el presente caso práctico se pretende encontrar un buen modelo en base al histórico de datos disponible.

Este caso práctico es una simplificación de la competición “*Walmart Recruiting - Store Sales Forecasting*” disponible en la comunidad de científicos de datos Kaggle. La diferencia entre el caso presentado y dicha competición es que esta última pretende predecir las ventas de 45 establecimientos diferentes. Debido al gran tiempo de computación necesario para ello y al poco valor añadido que presenta realizar las predicciones para los aproximadamente 100

departamentos que posee cada uno de los 45 establecimientos, se ha decidido estimar las ventas de los 98 departamentos de únicamente de uno de los establecimientos; en concreto, del establecimiento número cuatro.

El presente caso práctico ha sido resuelto utilizando el lenguaje de programación Python. Por tanto, todas las librerías y extractos de código presentados a continuación hacen referencia a dicho lenguaje.

A continuación, y con el objetivo de definir un modelo de previsión de ventas adecuado, se utiliza la metodología presentada en el apartado previo.

## **6.2. Precisión necesaria del modelo de previsión dependiendo de la tipología de producto**

Tal y como se ha mencionado al presentar el caso práctico, se dispone de las unidades vendidas en cada uno de los departamentos del establecimiento Walmart pero no se dispone del detalle de cuántas unidades han sido vendidas de cada producto.

En caso de que se dispusiera de las ventas de cada producto y debido a que los establecimientos Walmart ofertan una gran cantidad de productos, se debería considerar agrupar los productos por familias y generar conjuntamente las previsiones de productos que satisfacen una misma necesidad del consumidor.

Por otro lado, y en coherencia con las conclusiones extraídas en los apartados previos, sería conveniente realizar un análisis de agregación que permitiera distinguir qué familias de productos son críticas para la gestión de inventarios.

Para ello, se debería considerar que las familias de productos que generan más ventas y/o ingresos, que disponen de un tiempo de abastecimiento menor, que generan un mayor coste de stock (debido al espacio que ocupan o a necesidades especiales como la necesidad de refrigeración) o que disponen de un menor tiempo de frescura (fecha de caducidad cercana a la fecha de producción), se deben considerar como familias críticas. También se debería considerar como más críticas aquellas que debido a razones estratégicas son importantes para la compañía. En concordancia, sería oportuno considerar, por ejemplo, como más críticas aquellas en las que se desea aplicar una acción comercial para aumentar sus ventas.

Una vez divididas las familias de productos en críticas y no críticas, se debería aplicar la metodología presentada y utilizar métodos cuantitativos para obtener los modelos de previsión de ventas. Se realizaría el diseño de los modelos de previsión por separado, por un lado, se analizarían las familias críticas y por el otro lado las no críticas. Para diseñar las previsiones

de las familias críticas se debería invertir en el tiempo de diseño y establecer cuáles son los mejores parámetros y modelos que permiten unas mejores previsiones; contemplándose la posibilidad de disponer de diferentes modelos de previsión para las diferentes familias críticas con el fin de obtener una precisión de las previsiones de venta superior al 90%. Por otro lado, se generaría un modelo conjunto de las familias de productos no críticas en el que se invertiría menor tiempo de diseño y en el que se aceptaría una precisión del 80%.

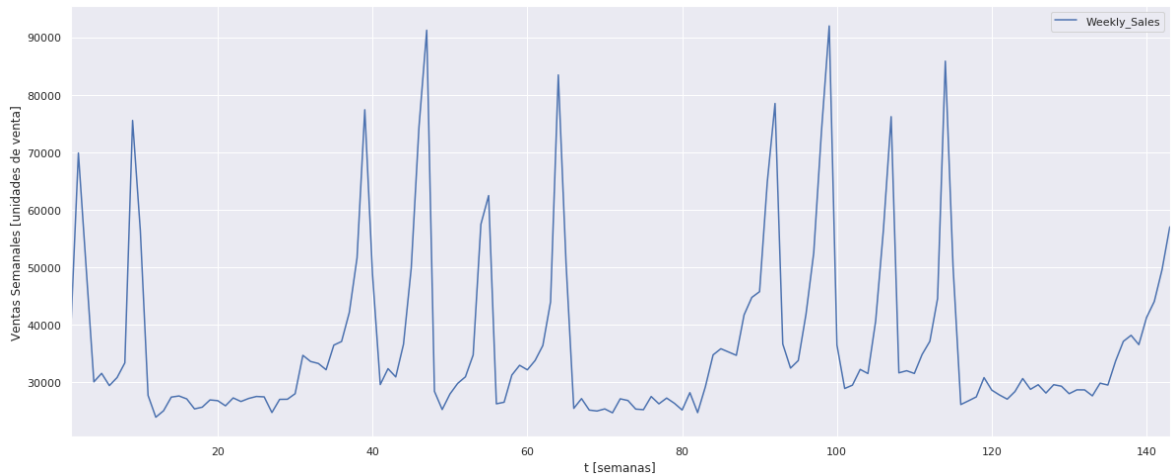
Respecto a las previsiones de productos nuevos, se debe destacar que se considerarían de igual forma que el resto de productos. En caso de que estos artículos formasen parte de una familia de productos existentes, ninguna consideración especial sería necesaria; ya que se dispondría de históricos de datos de la familia. En caso de introducirse una nueva familia de artículos, se deberían utilizar datos históricos de una familia de productos que satisfaga una necesidad similar del consumidor o considerar como histórico de datos las ventas medias generadas en el establecimiento al introducir nuevas familias.

Cabe destacar, que la metodología presentada considera establecer criterios cualitativos para valorar los resultados de las previsiones generadas por los modelos cuantitativos. En el caso de valorar nuevas familias de productos, sería apropiado invertir mayores recursos en esta etapa.

Debido a que no se dispone de las unidades vendidas de cada producto y con el fin de evaluar la posibilidad de obtener modelos precisos, en los apartados consecutivos se desea obtener un modelo de previsión altamente preciso para las ventas de cada uno de los departamentos del establecimiento Walmart considerado. Dicha precisión será necesaria para las familias de productos críticos, pero no para la totalidad de los artículos.

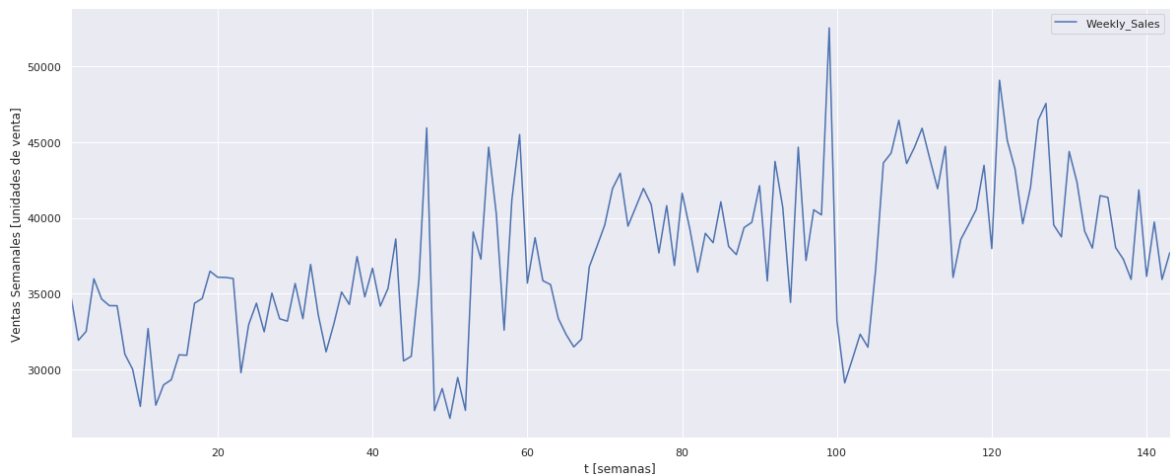
### **6.3. Exploración y manipulación de los datos**

Con el fin de conocer y explorar los datos a tratar en el presente caso práctico, se realizan representaciones gráficas de las unidades de venta semanales para los distintos departamentos del establecimiento Walmart considerado. En la Figura 6.2 y en la Figura 6.3 se representan, respectivamente, las ventas semanales registradas en el histórico de datos para el departamento 1 y para el 10. Estos departamentos serán los utilizados para representar gráficamente las previsiones de los modelos candidatos considerados en los apartados consecutivos.



*Figura 6.2 Ventas semanales del departamento 1 del establecimiento Walmart considerado*

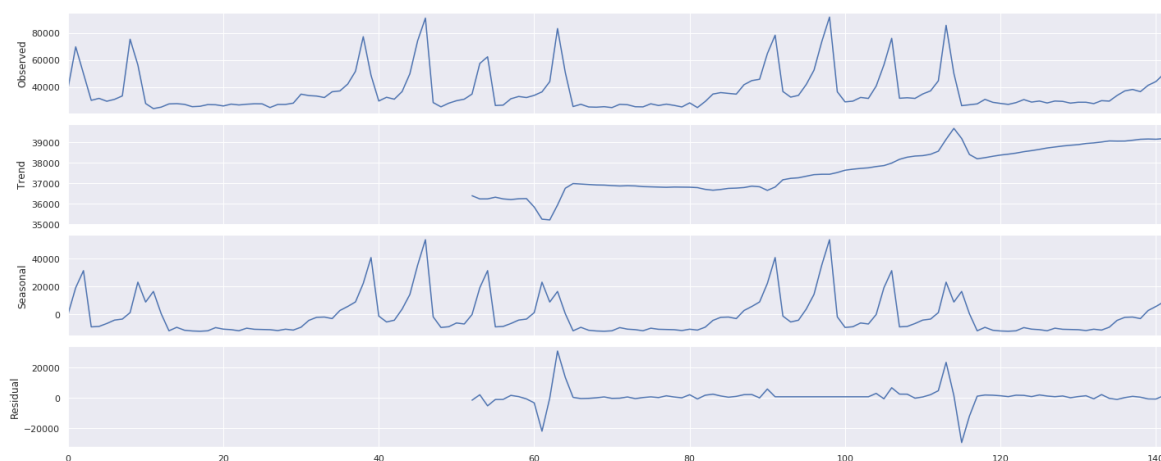
*Fuente: Propia*



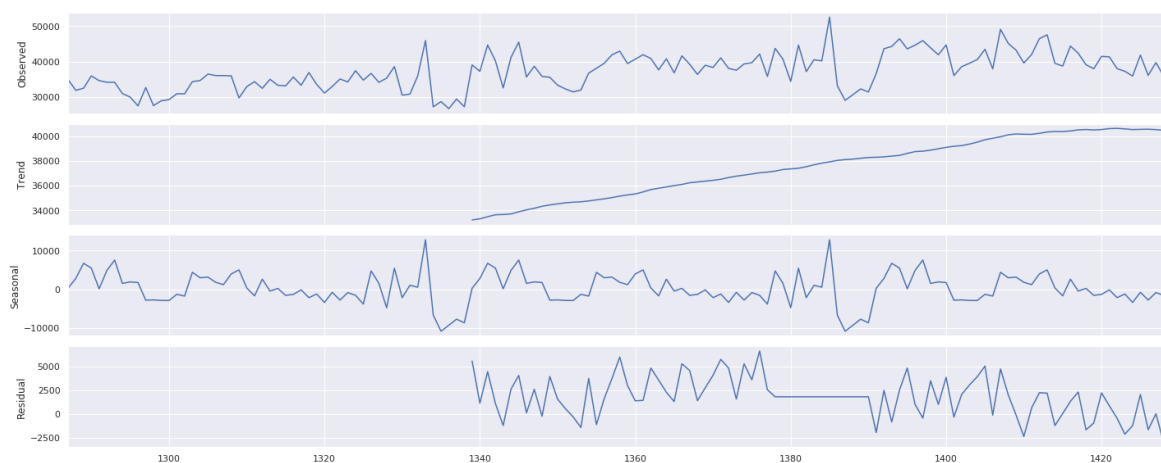
*Figura 6.3 Ventas semanales del departamento 10 del establecimiento Walmart considerado*

*Fuente: Propia*

La inspección visual de las figuras anteriores permite observar que existe una variación estacional en los datos de ventas. En la Figura 6.4 y en la Figura 6.5 se representa la descomposición estacional de los datos en tendencia y estacionalidad aditiva. Tal y como puede observarse, la estacionalidad es constante de un año a otro para cada semana del año. Por ese motivo, se decide que a la hora de considerar modelos candidatos de previsión de causalidad, se predecirá la tendencia-ciclo de las ventas semanales y se asumirá que la variación estacional es un dato conocido. Por otro lado, al considerar modelos de series de tiempo, debido a la naturaleza de estos, se considerarán los datos registrados de las unidades de venta.



*Figura 6.4 Descomposición estacional de las ventas del departamento 1 Fuente: Propia*



*Figura 6.5 Descomposición estacional de las ventas del departamento 10 Fuente: Propia*

A pesar de que en el caso de los departamentos 1 y 10 la tendencia de los datos es creciente, no es así en todos. Por ejemplo, en el departamento 6, representado en la Figura 6.1, la tendencia de los datos decrece en el tiempo.

Además, con el objetivo de seleccionar correctamente las variables explicativas a formar parte de los posibles modelos candidatos, aquellas variables que permiten definir la variable objetivo de las ventas o de la tendencia de estas; se realiza un análisis de las correlaciones entre variables, los resultados del cual se presentan en la Figura 6.6.



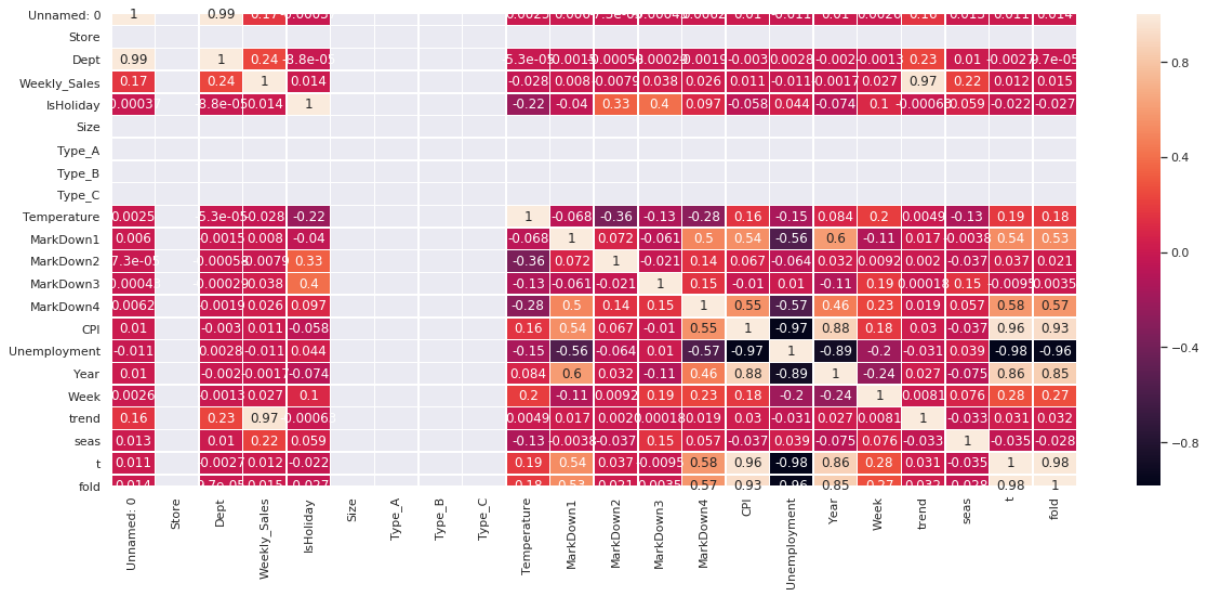


Figura 6.6 Análisis de las correlaciones entre variables Fuente: Propia

En la figura anterior aparecen las variables de identificación del establecimiento y de la dimensión y tipología de este, valores de las cuales son constantes en el caso práctico considerado, ya que únicamente se considera un establecimiento. Por tanto, estas no serán incluidas en los diversos modelos. Por otro lado, se observa que existen variables altamente correlacionadas entre sí; valores del coeficiente de autocorrelación próximos a la unidad, ya sea positiva o negativa. En concreto, la variable de la tasa de desempleo (*unemployment*) está altamente relacionada con el CPI y con variables temporales. Con el fin de no incluir en el modelo variables que intentan explicar el mismo fenómeno, se excluye también la variable de la tasa de desempleo. Por último, en vez de utilizar el registro de fechas temporales disponible en la base de datos (año, mes y día), se decide utilizar la variable temporal *t*; la cual registra el número de la semana, considerado como semana 0 aquella en la que se inicia la recogida de datos.

Recaltar que, en el presente caso práctico se considera como variable de interés y las unidades de venta semanales, o dependiendo del modelo los datos de la tendencia, y como variables explicativas *X* el resto de variables consideradas una vez los datos han sido manipulados: departamento (*Dept*), si es festivo (*IsHoliday*), temperatura (*Temperature*), *MarkDown1*, *MarkDown2*, *MarkDown3*, *MarkDown4*, *CPI* y *t*. Al aplicar modelos de series de tiempo se predecirá las unidades de venta. Por otro lado, a la hora de aplicar modelos de previsión causales de aprendizaje automático, los algoritmos en que estos se basan predecirán la tendencia de las ventas. A continuación, a dichos valores se sumarán los datos de estacionalidad y estos serán los datos considerados para valorar el rendimiento de dichos modelos.

## 6.4. División de los datos para la valoración de los modelos de previsión candidatos

Para resolver el presente caso práctico se han considerado un 80% de los datos históricos disponibles como datos de entrenamiento y un 20% como datos de prueba. De esta forma, a excepción del modelo de serie temporal básico considerado donde datos de entrenamiento no son necesarios, un 80% de los datos se utilizan para entrenar los algoritmos y formar el modelo y el 20% de datos restantes se utiliza para valorar la precisión del modelo dadas entradas desconocidas.

En concreto, se valora la precisión de los distintos modelos considerados calculando el error relativo de las previsiones de ventas, dadas las entradas de los datos de prueba, respecto el valor real de dicha variable.

## 6.5. Aplicación de diversos modelos candidatos y valoración de los resultados

A continuación se consideran diversos modelos candidatos de los analizados en la sección 4.

Se debe destacar que en el presente caso se considera que se puede estimar la demanda estimando las ventas; es decir, se supone que en el pasado toda la demanda fue abastecida. En caso de que esto no fuera así, y se dispusiera de datos sobre cuándo la demanda no ha sido abastecida, se podría predecir la demanda en vez de las ventas.

### 6.5.1. Modelo simple de medias móviles

En primer lugar, se considera el modelo simple de medias móviles, modelo basado en el análisis de series de tiempo. Debido a la naturaleza del modelo, se decide predecir con él las unidades de venta y no la tendencia de estas. Además, únicamente pueden considerarse los datos históricos de la variable de interés  $y$ , siendo esta las unidades de venta, y no se considera ninguna variable explicativa  $X$ . Tampoco es necesario considerar la división de los datos en datos de entrenamiento y datos de prueba.

El modelo es aplicado en cada una de los departamentos del establecimiento por separado y se considera un orden del modelo igual a 4, es decir se aplica el modelo SMA (4). Para la aplicación del modelo se utilizan la función de Python [31]:

$$pred = y.rolling(window = 4).mean() \quad (Ec. 6.1)$$

Esta permite realizar las predicciones de la variable de interés ( $pred$ ) como media de los 4 últimos valores de dicha variable ( $y$ ). En la Figura 6.7 y en la Figura 6.8 se muestra la representación gráfica de aplicar este modelo en los departamentos 1 y 10.

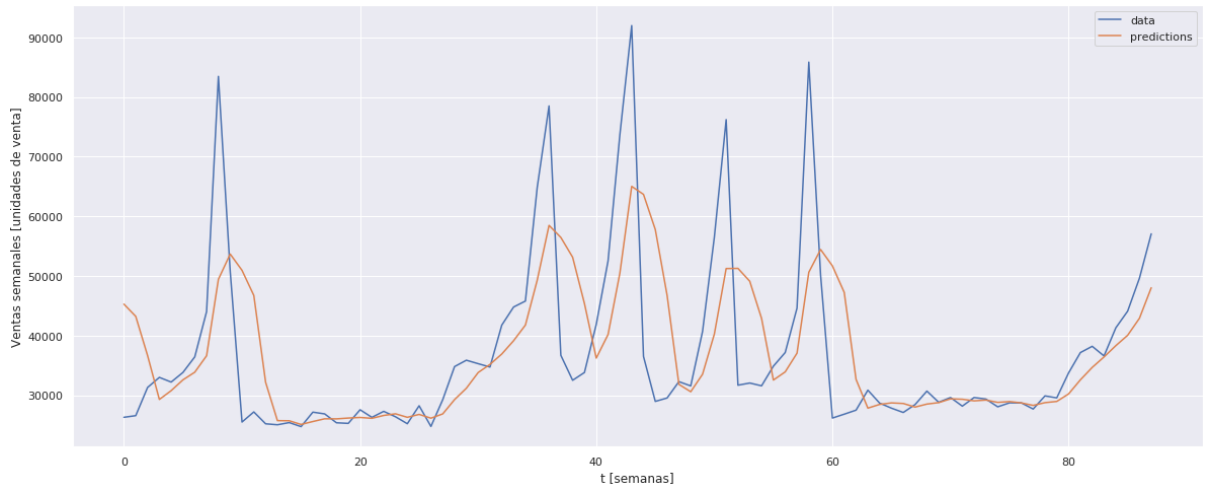


Figura 6.7 Aplicación del modelo SMA (4) en el departamento 1 Fuente: Propia

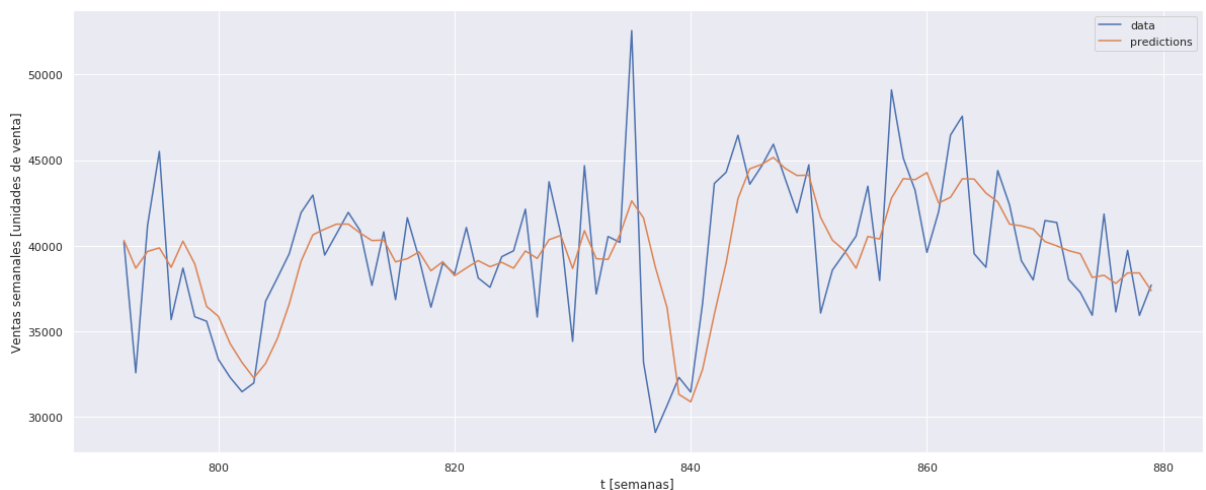


Figura 6.8 Aplicación del modelo SMA (4) en el departamento 10 Fuente: Propia

Considerando todos los departamentos del establecimiento, el error relativo de las previsiones del modelo SMA es del 114,85%.

### 6.5.2. Modelo estacional autorregresivo integrado de medias móviles con entrada exógena

A continuación, se considera como modelo candidato el modelo estacional autorregresivo de medias móviles con entrada exógena (SARIMAX). Debido a la naturaleza del modelo y a que este considera la variación estacional de los datos, el modelo se aplica sobre los datos de las

unidades de venta, y no sobre la tendencia de estos. Este modelo también es aplicado a cada departamento por separado. Por tanto, se considera como variable de interés y las unidades de venta y todas las variables explicativas  $X$  consideradas, a excepción de la variable departamento, ya que al realizarse el modelo por departamento tendrá valor constante.

Como se ha visto al analizarse los modelos de series de tiempo, el modelo SARIMAX incluye al igual que el modelo SARIMA 7 parámetros:  $(p, d, q) \times (P, D, Q)_m$ . Con el fin de escoger los valores de los parámetros que permiten obtener un modelo de previsión lo más preciso posible; en primer lugar, se analizan los datos para definir el valor adecuado de los parámetros o para conocer qué posibles valores pueden ser adecuados dados los datos del departamento uno. En segundo lugar, en base a los resultados obtenidos de dicho análisis, se utiliza la técnica *Grid Search* con el fin de establecer qué valores próximos a los encontrados para el departamento 1 se consideran los más adecuados para cada uno de los departamentos.

Primero, se analizan los valores de los parámetros de la parte del modelo no estacional, los parámetros  $(p, d, q)$ . Para definir estos valores, se analizan los datos de la tendencia, es decir no se consideran las variaciones introducidas por la estacionalidad. Para analizar el valor adecuado para el parámetro  $d$  se examina cuántas veces deben ser diferenciados los datos para eliminar la tendencia. Tal y como se observa en la Figura 6.9, diferenciado los datos una vez la tendencia se elimina; por tanto, puede considerarse  $d = 1$ .

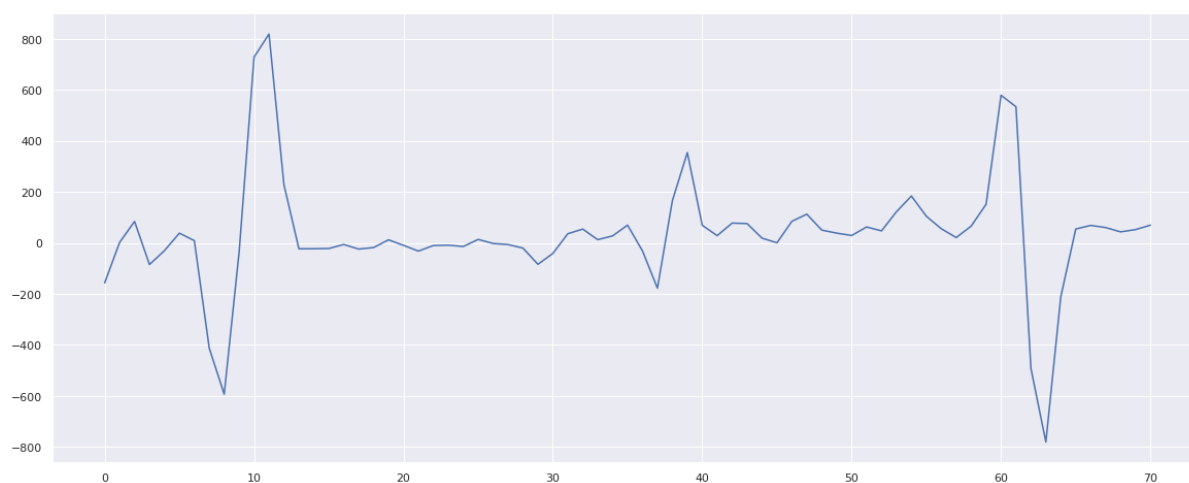


Figura 6.9 Datos de la tendencia de las unidades vendidas en el departamento 1 diferenciados una vez Fuente: Propia

Con el fin de conocer los valores adecuados para los parámetros  $p$  y  $q$ , se analizan los gráficos de autocorrelación parcial (PACF) y de autocorrelación (ACF) para los datos de la tendencia

[32]. Estos gráficos se representan en la Figura 6.10. El parámetro  $p$  puede ser escogido observando el número de correlaciones significativas en el gráfico PACF antes de poderse considerar despreciables; el parámetro  $q$  puede ser escogido utilizando el mismo criterio pero analizando el gráfico ACF. Por tanto, en este caso, se pueden considerar como valores adecuados  $p = 3$  y  $q = 2$ .

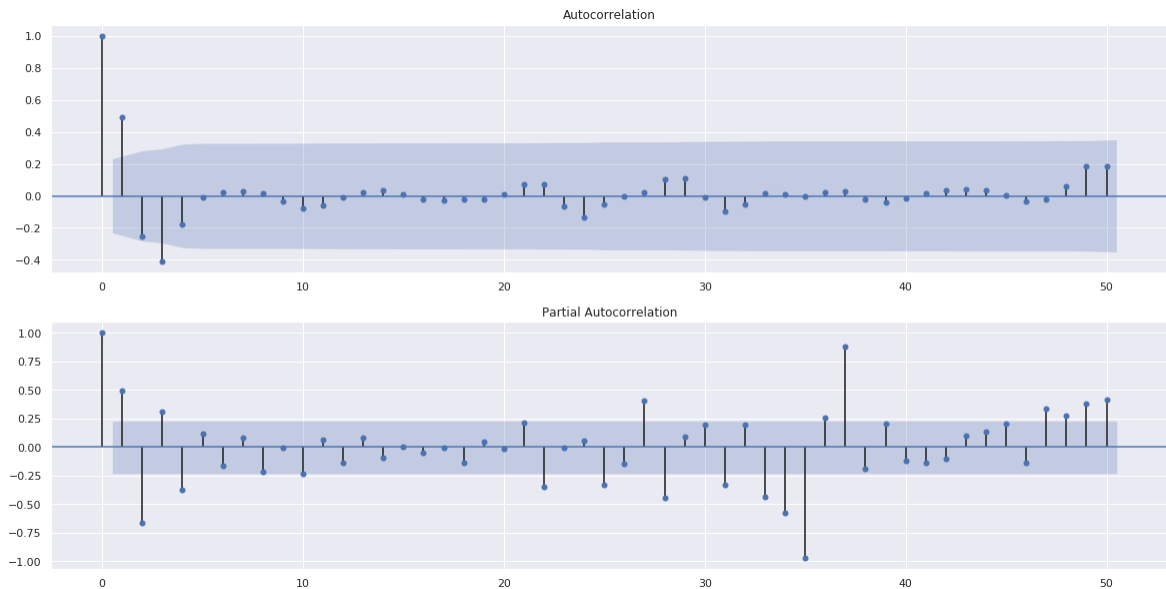


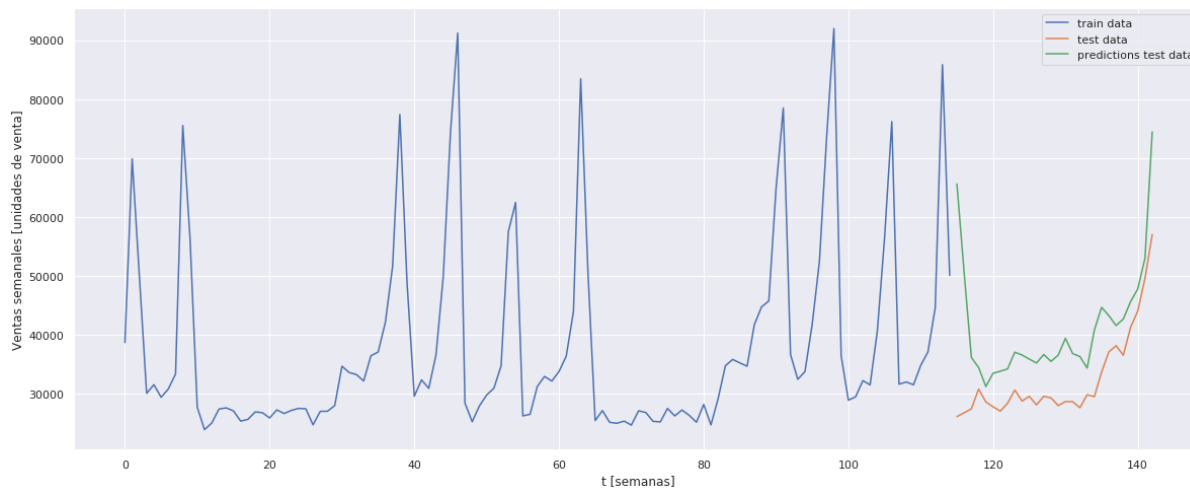
Figura 6.10 Gráficos ACF y PACF para los datos de la tendencia del departamento 1

Fuente: Propia

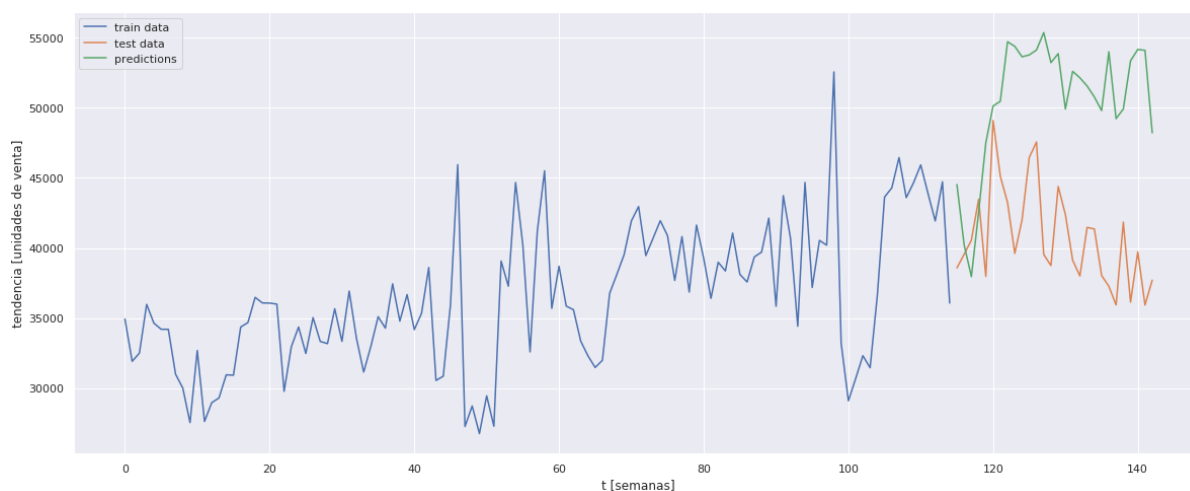
Respecto a los valores de los parámetros de la parte estacional del modelo  $(P, D, Q)_m$ , se considera que el parámetro  $m$  es 52, ya que se dispone de los datos semanales; es decir, de 52 valores al año de la variable de interés. Por otro lado, los valores de los parámetros  $(P, D, Q)$  se analizan de forma análoga a los parámetros  $(p, d, q)$ , pero considerando la parte estacional de los datos. Como resultado del análisis se obtiene que serían valores adecuados  $P = 1$ ,  $D = 1$  y  $Q = 0$ .

Tal y como se ha presentado con anterioridad, estos valores mencionados para los distintos parámetros, se han encontrado analizando los datos del departamento 1. Por tanto, es probable, que los valores más adecuados para otros departamentos sean similares pero varíen; por ello, se decide aplicar la técnica *Grid Search* para cada uno de los departamentos considerado para cada parámetro tres valores, a excepción de para el parámetro  $m$  valor el cual será invariable. En la mayoría de los casos, los tres valores contemplados son el valor encontrado del análisis del departamento 1, la unidad superior y la inferior. Por tanto, mientras que para el departamento 1 se han definido los valores  $(3, 1, 2) \times (1, 1, 0)_{52}$ , el método *Grid Search* considera que, por ejemplo, para el departamento 10 los valores adecuados son  $(2, 1, 2) \times (1, 1, 0)_{52}$ .

Aplicando los valores de los parámetros resultado del análisis presentado en la función `tsa.statespace.sarimax.SARIMAX` de la librería de Python statsmodel [33], se obtienen las previsiones representadas en la Figura 6.11 y en la Figura 6.12.



*Figura 6.11 Previsiones de ventas del departamento 1 al aplicar modelo SARIMAX Fuente: Propia*



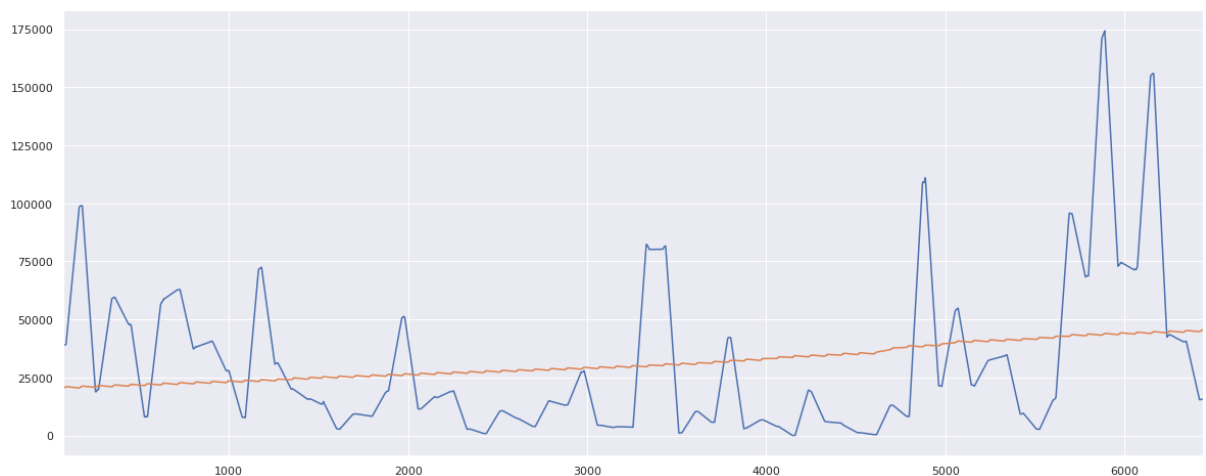
*Figura 6.12 Previsiones de ventas del departamento 10 al aplicar modelo SARIMAX Fuente: Propia*

El error relativo de las previsiones de venta, considerando todos los departamentos del establecimiento Walmart, realizadas por el modelo SARIMAX dadas las entradas  $X$  de los datos de prueba es del 31,29 %.

### 6.5.3. Modelo de regresión lineal

En esta sección se implementa el modelo de aprendizaje automático de regresión lineal. Al igual que en los demás modelos basados en el aprendizaje automático que serán considerados, se realiza las previsiones de la tendencia de las unidades de venta y la estacionalidad se considera como un dato conocido.

Debido a que realizar un modelo único considerando la totalidad de los datos y todas las variables no presenta unos resultados adecuados, tal y como se observa en la Figura 6.13, se realiza un modelo de regresión lineal para cada uno de los departamentos. En la Figura 6.13 y en la Figura 6.14 se representan, en azul, los datos de las tendencias de todos los departamentos del establecimiento Walmart y en rojo, las previsiones de estos datos realizados por el modelo de regresión lineal global y el modelo realizado por cada tienda. En el modelo de regresión global, representado en la primera de las figuras, se consideran todas las variables  $X$  y la totalidad de los datos conjuntamente. Por otro lado, en el modelo de regresión realizado por departamento, se aplica el modelo a cada departamento por separado y se elimina de las  $X$  la variable del departamento ya que se convierte en una constante. Dichas figuras evidencian la necesidad de aplicar el modelo de regresión lineal por departamento.



*Figura 6.13 Aplicación del modelo de regresión lineal global considerando todos los datos de las tendencias de cada departamento conjuntamente Fuente: Propia*

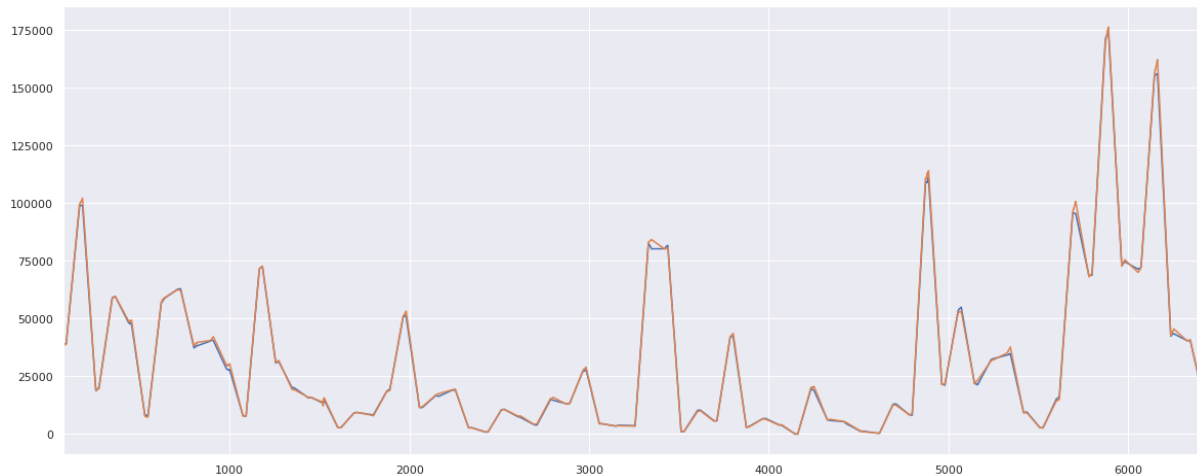


Figura 6.14 Aplicación del modelo de regresión lineal sobre los datos de las tendencias de cada departamento por separado Fuente: Propia

Por tanto, se aplica el modelo de regresión lineal en cada departamento y se predice la tendencia de las unidades de venta sin considerar la variable departamento como parte de las X. Para ello, se utiliza el algoritmo *LinearRegression* de la librería de Python *sklearn.linear\_model* [34] y se entrena dicho algoritmo con los datos de entrenamiento. Las previsiones de la tendencia de las ventas para los departamentos 1 y 10 dadas las entradas de los datos de prueba se representan a continuación.

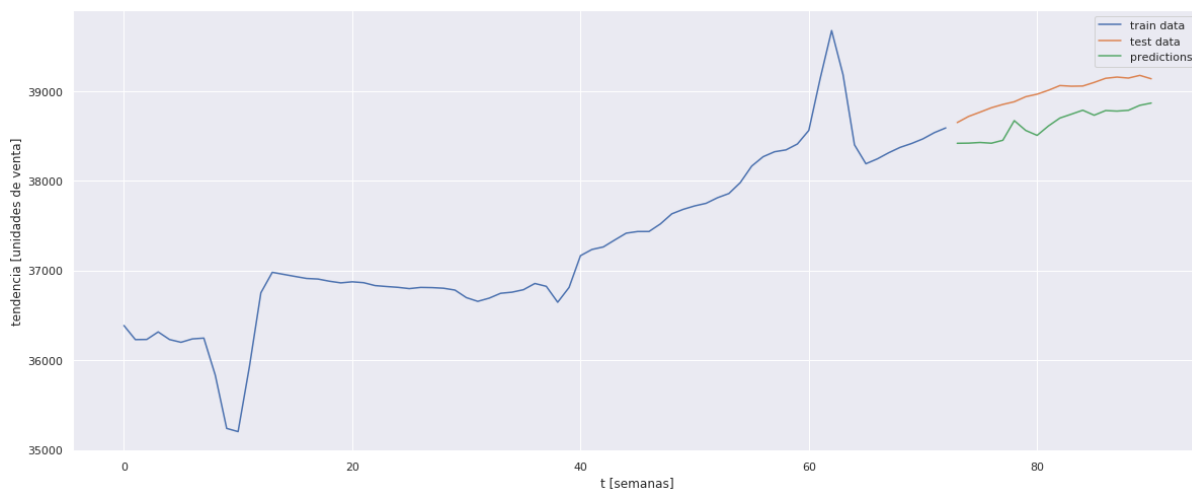
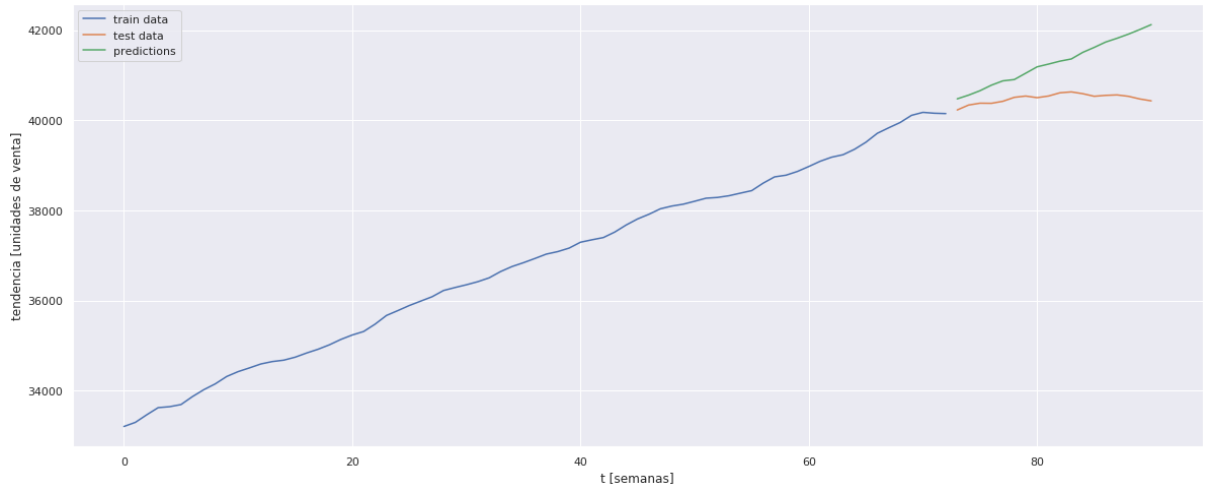


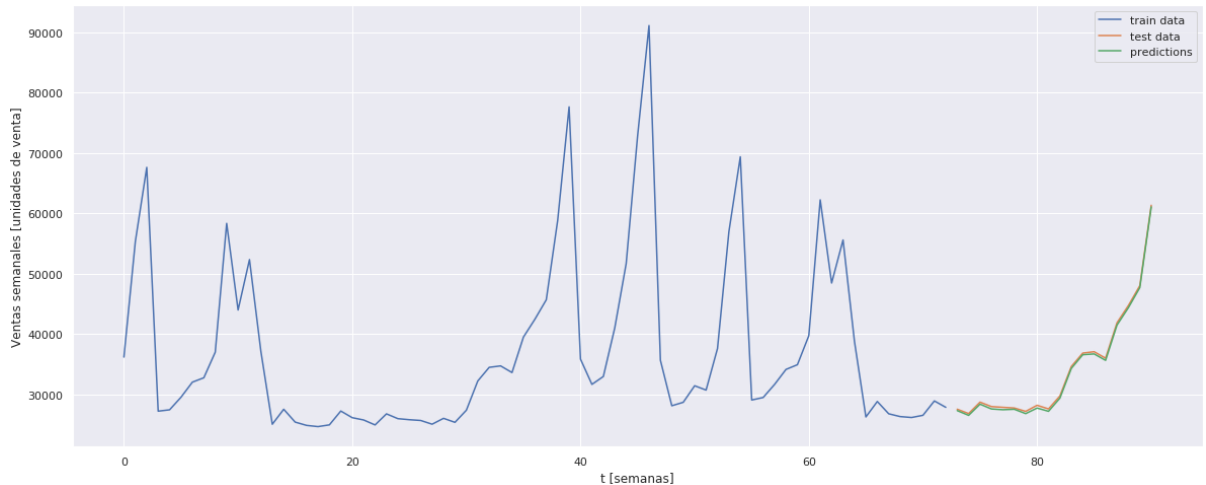
Figura 6.15 Previsiones de la tendencia de las ventas del departamento 1 al aplicar modelo de regresión lineal Fuente: Propia





*Figura 6.16 Previsiones de la tendencia de las ventas del departamento 10 al aplicar modelo de regresión lineal Fuente: Propia*

Una vez conocidas las previsiones de la tendencia se suman a ellas los datos de la estacionalidad, considerado como un valor conocido para cada departamento y semana del año. Como resultado se obtiene el valor de la variable real de interés, que son las ventas semanales para cada departamento. El resultado gráfico de las previsiones de dicha variable se presenta a continuación.



*Figura 6.17 Previsiones de ventas del departamento 1 al aplicar modelo de regresión lineal Fuente: Propia*

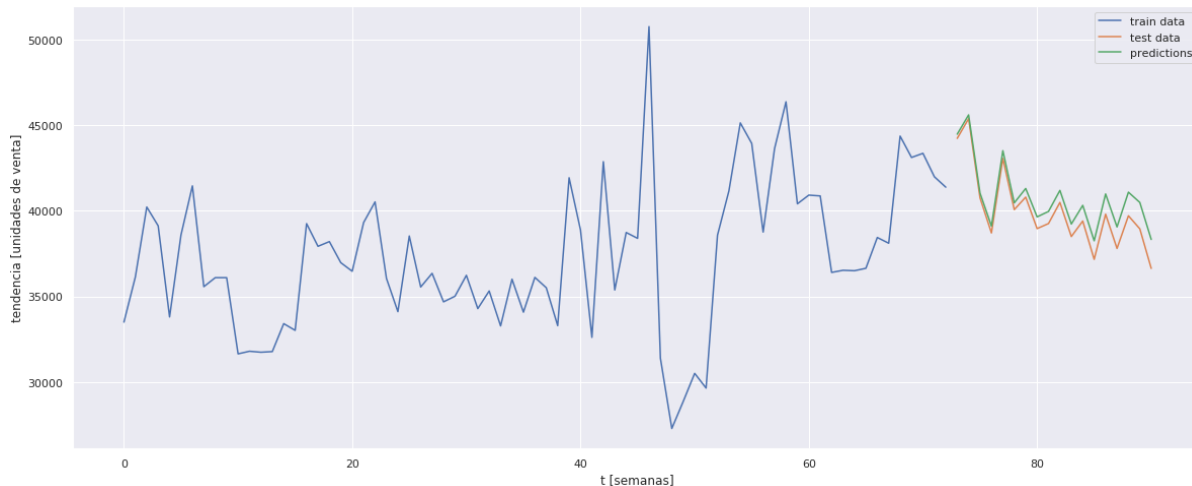


Figura 6.18 Previsiones de ventas del departamento 10 al aplicar modelo de regresión lineal

Fuente: Propia

El error relativo de las previsiones de las ventas proporcionado por el modelo de regresión lineal es del 7,06 %.

#### 6.5.4. Modelo de k-vecinos más cercanos

El siguiente modelo candidato considerado es el basado en el algoritmo de k-vecinos más cercanos adaptado para realizar regresiones. Al igual que en el caso de la regresión lineal, el modelo debe ser aplicado por cada departamento por separado. Además, se considera también que la variable de interés y a predecir con el modelo de aprendizaje automático es la tendencia de las unidades de venta y de las variables  $X$  consideradas se elimina la variable departamento.

Debido a la naturaleza del modelo, y con el fin de que las medidas de similitud que este realiza sobre los diferentes puntos pertenecientes a  $X$  sean coherentes, se escalan los diferentes valores de  $X$  previamente a entrenar los algoritmos del modelo. Para ello se utiliza la función *StandardScaler* de la librería de Python *preprocessing* [35].

Con el fin de encontrar un modelo basado en el algoritmo de k-vecinos más cercanos adecuado, se utiliza la técnica *Grid Search* para seleccionar los valores de los parámetros del modelo para cada departamento. El algoritmo utilizado para entrenar el modelo es el *neighbors.KNeighborsRegressor* de la librería de Python *sklearn* [36] y para realizar la búsqueda de cuadrículas se utiliza la función *model\_selection.GridSearch* de la misma librería. Los parámetros considerados en la búsqueda de cuadrículas son:

- El número de vecinos: se consideran diversos valores discretos entre 2 y 35.
- El orden de la distancia de Minkowski: esta es la medida de similitud considerada por el algoritmo y los diferentes valores del orden que se contemplan son 1 y 2.
- Los pesos utilizados para realizar las ponderaciones de las predicciones dados los valores de los distintos vecinos: se considera que estos pesos pueden ser uniformes (cada vecino pondera lo mismo) o basados en la distancia (los vecinos más cercanos ponderan más).
- El algoritmo utilizado para encontrar los vecinos más cercanos.

Los valores de los parámetros resultado de la búsqueda de cuadrículas varían por departamento, de todas formas, en la mayoría de los casos se consideran cinco vecinos y una ponderación para realizar las predicciones basada en la distancia.

Como resultado de aplicar lo previamente descrito, se encuentran las previsiones de la tendencia de las unidades vendidas representadas en la Figura 6.19 y en la Figura 6.20 para los departamentos uno y diez.

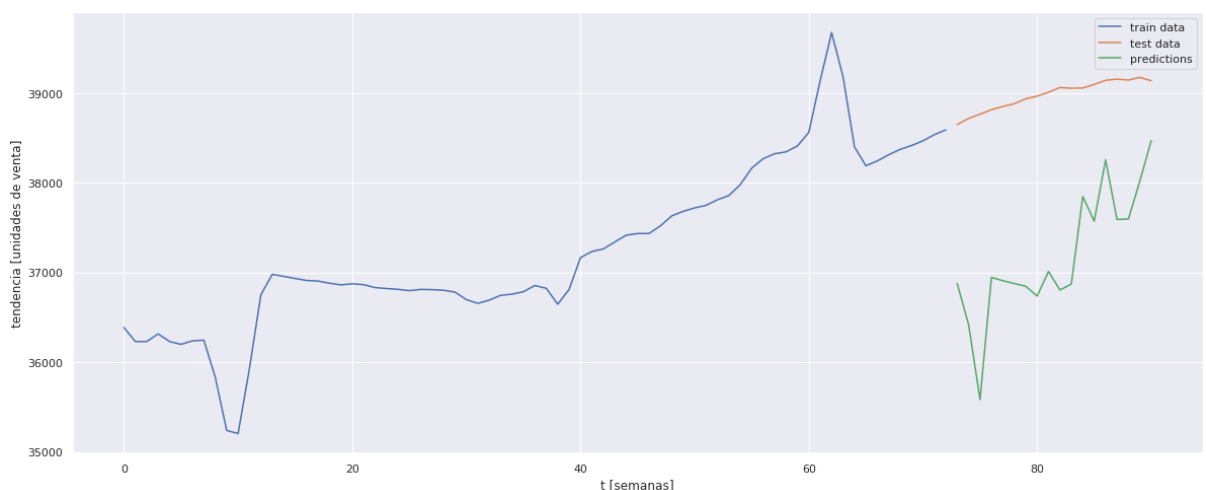


Figura 6.19 Previsiones de la tendencia de las ventas del departamento 1 al aplicar modelo de k-vecinos más cercanos Fuente: Propia

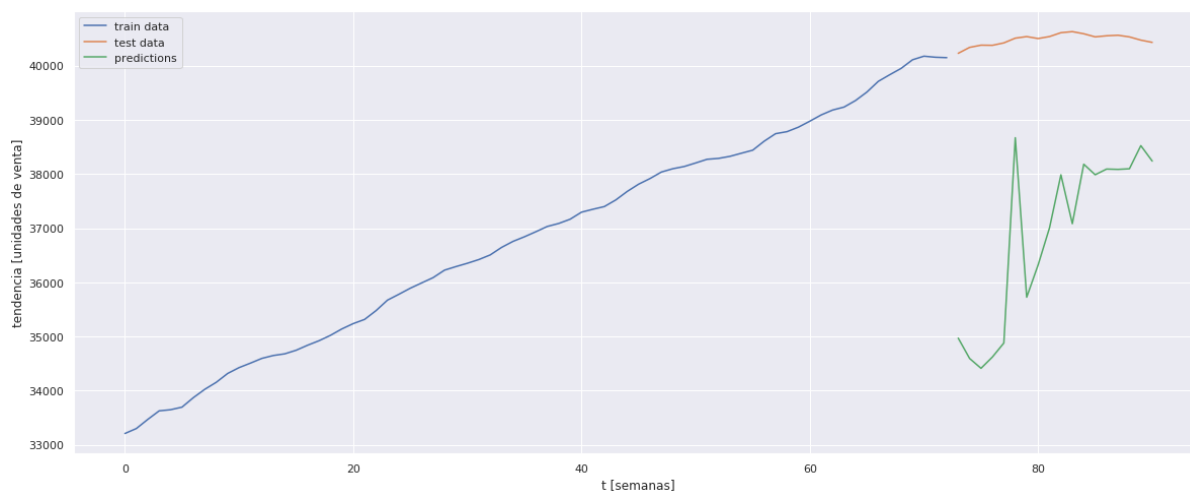


Figura 6.20 Previsiones de la tendencia de las ventas del departamento 10 al aplicar modelo de  $k$ -vecinos más cercanos Fuente: Propia

Sumando los datos de la estacionalidad se encuentran las previsiones de las unidades de venta, las cuales se representan en la Figura 6.21 y en la Figura 6.22. El error relativo de este modelo para predecir las unidades de venta de los datos de prueba es del 16,85 %.

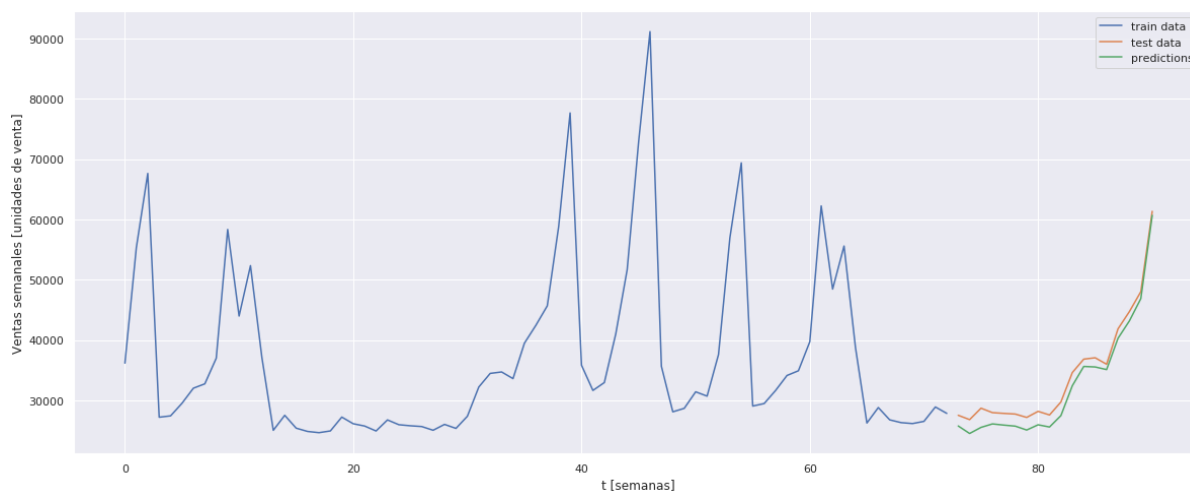


Figura 6.21 Previsiones de ventas del departamento 1 al aplicar modelo de  $k$ -vecinos más cercanos Fuente: Propia

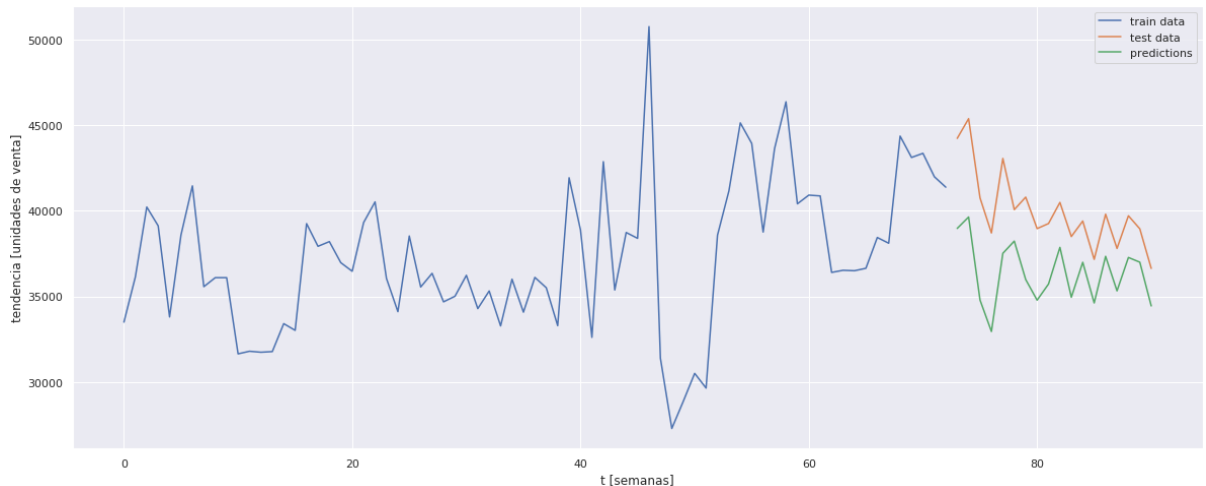


Figura 6.22 Previsiones de ventas del departamento 10 al aplicar modelo de *k*-vecinos más cercanos

Fuente: Propia

### 6.5.5. Modelo de bosques aleatorios

A continuación, se presenta el modelo de bosques aleatorios. Para este modelo se considera también como variable a predecir por el modelo de aprendizaje automático la variable de la tendencia de las unidades de venta, a posteriori los datos conocidos de la estacionalidad serán sumados a las predicciones encontradas para obtener así las predicciones de las unidades de venta.

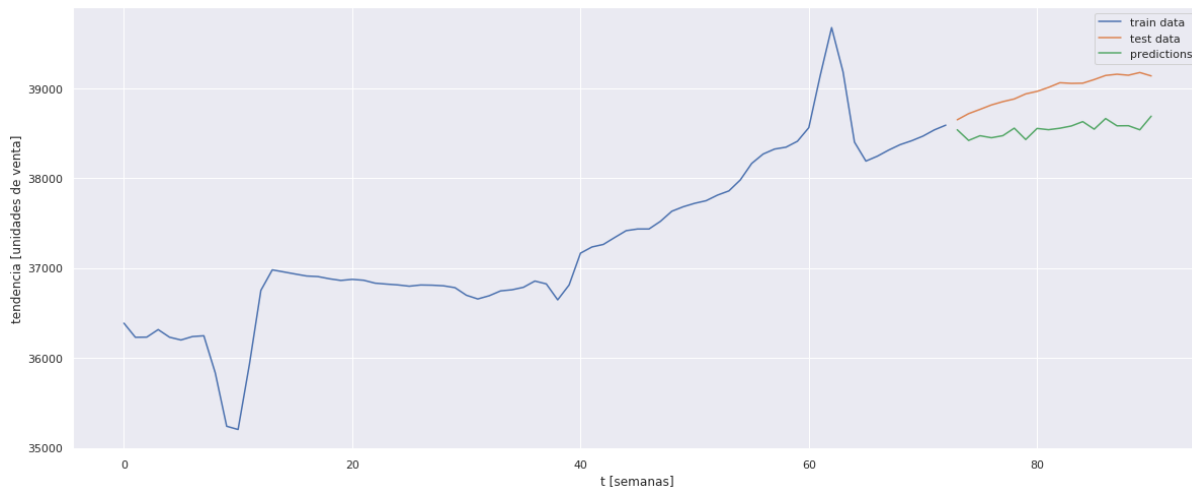
El algoritmo de bosques aleatorios considerado es el *ensemble.RandomForestRegressor* de la librería *sklearn* de Python [37]. Dicho algoritmo permite crear un modelo fiable considerando todos los datos del histórico en conjunto y no considerando los de cada uno de los departamentos por separado.

Se realiza también una búsqueda de cuadrículas para encontrar aquellos valores de los parámetros del algoritmo que permiten obtener unos resultados más precisos. Algunos de los parámetros considerados para la búsqueda y de los resultados obtenidos son:

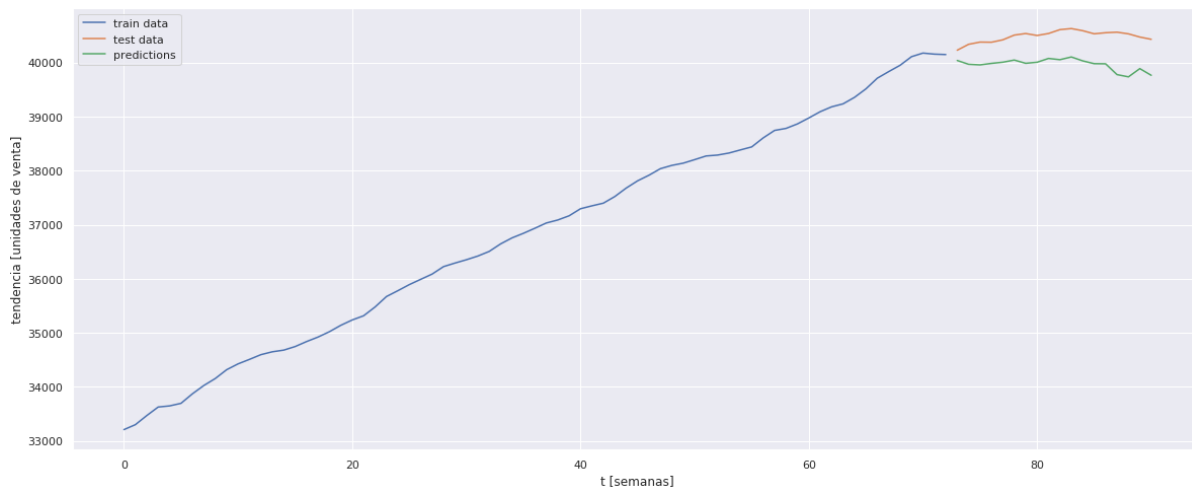
- El número de árboles considerados en el bosque. Se consideran diversos valores entre 100 y 2000 y como resultado de la búsqueda el más adecuado es 200.
- *Bootstrap*: si cierto, se consideran muestras de los datos de entrenamiento para construir cada uno de los árboles; si falso, se utilizan la totalidad de los datos para construir cada árbol. Como resultado, se obtiene que el valor más adecuado es que sea falso.
- La profundidad máxima de cada árbol. Dicha profundidad puede ser limitada o no. Se contemplan ambas opciones en la búsqueda de cuadrículas y se obtiene que no limitarla proporciona mejores resultados.

- El número de atributos considerados para realizar cada subdivisión de datos. Este valor puede ser un número concreto o función del número total de atributos. Se consideran ambas opciones y se obtiene que lo más adecuado es considerar como máximo tantos atributos como la raíz cuadrada del total de atributos.

Como resultado de aplicar el algoritmo regresivo de bosques aleatorios con los valores de los parámetros mencionados, se encuentran las previsiones de la tendencia de las unidades de venta semanales representadas a continuación.

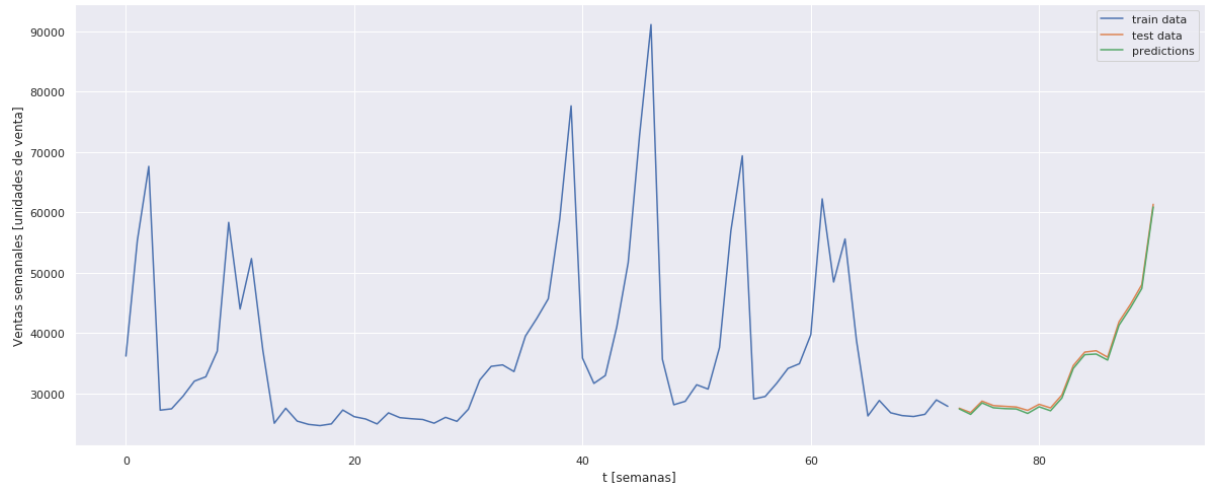


*Figura 6.23 Previsiones de la tendencia de las ventas del departamento 1 al aplicar modelo de bosques aleatorios Fuente: Propia*



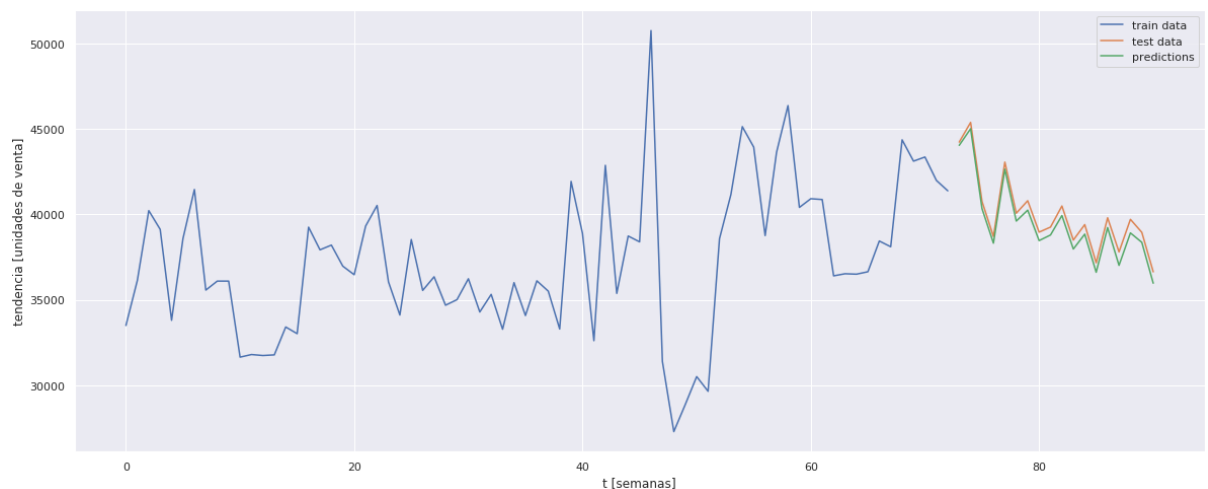
*Figura 6.24 Previsiones de la tendencia de las ventas del departamento 10 al aplicar modelo de bosques aleatorios Fuente: Propia*

Si a ello se le añade los datos de la estacionalidad aditiva, se pueden predecir las ventas dadas las entradas de los datos de prueba, tal y como se muestra en la Figura 6.25 y en la Figura 6.26. El error relativo de estas previsiones es del 3,20 %.



*Figura 6.25 Previsiones de ventas del departamento 1 al aplicar modelo de bosques aleatorios*

*Fuente: Propia*



*Figura 6.26 Previsiones de ventas del departamento 10 al aplicar modelo de bosques aleatorios*

*Fuente: Propia*

### 6.5.6. Modelo de redes neuronales

Se considera también como modelo candidato un modelo basado en redes neuronales. Este modelo se construye por departamento y consta en todos los casos de 8 neuronas en la capa de entrada, ya que se consideran 8 atributos de  $X$  (todos menos la variable departamento), 9 capas intermedias y una capa de salida con una sola neurona, ya que únicamente hay una variable a prever. La red neuronal se define gracias a la librería de Python *keras* [38] y la

variable que se desea predecir es la tendencia de las unidades de venta.

Con el fin de encontrar un modelo de redes neuronales adecuado, en la etapa de diseño, se contemplan diferentes números de capas intermedias, diferentes ponderaciones de la importancia de cada neurona y diferentes funciones de activación. Para definir unos valores adecuados de estos parámetros se utiliza la técnica de la búsqueda de cuadrículas.

Como resultado, se define la red neuronal con 9 capas intermedias y se utiliza la función de activación *Rectified Linear Unit* (ReLU). También se decide utilizar el optimizador Adam (*Adaptative Moment Estimation*) [39], el cual permitirá modificar el valor de parámetros de la red neuronal con el fin de aumentar la precisión de las previsiones; uno de los parámetros que evaluará será el peso de las ponderaciones de cada neurona. Este optimizador realiza una búsqueda exhaustiva de los valores de los parámetros cuando estos se aproximan al óptimo.

Las previsiones de los datos de la tendencia que realiza dicha red neuronal en los departamentos 1 y 10 se representan en la Figura 6.27 y en la Figura 6.28.

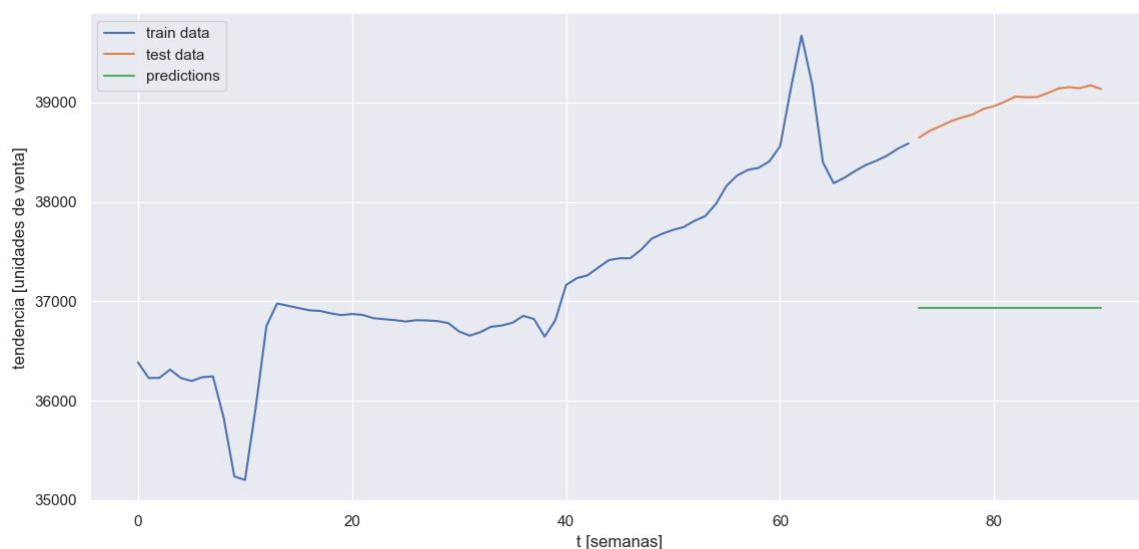
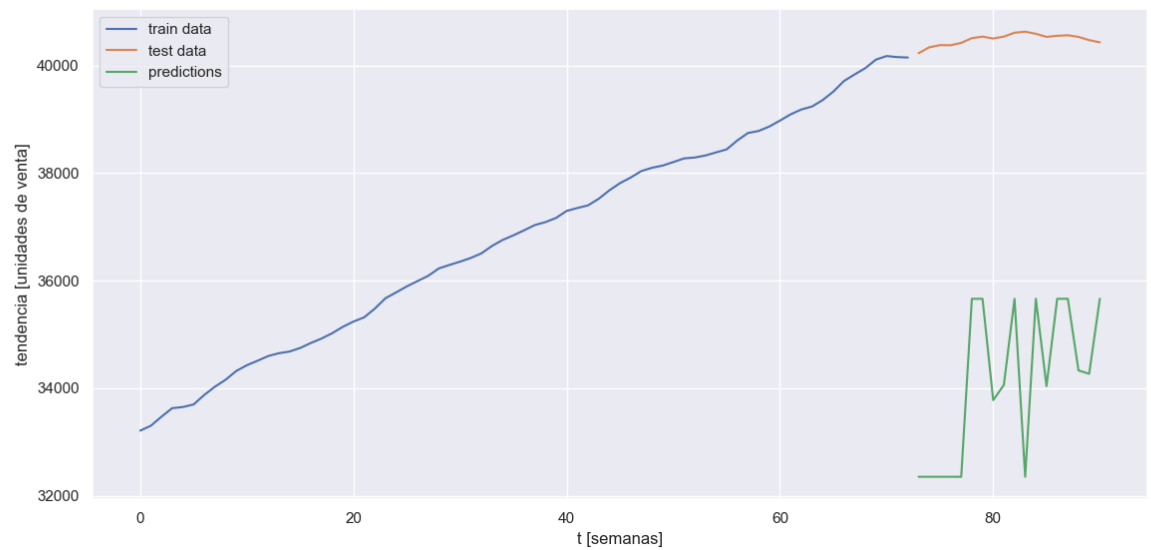


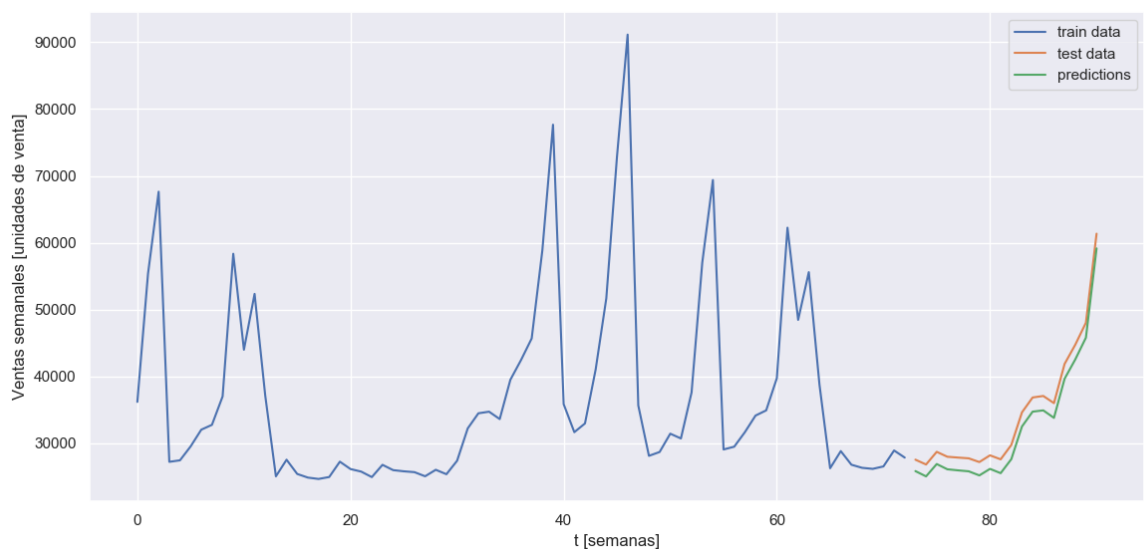
Figura 6.27 Previsiones de la tendencia de las ventas del departamento 1 al aplicar modelo de redes neuronales Fuente: Propia





*Figura 6.28 Previsiones de la tendencia de las ventas del departamento 10 al aplicar modelo de redes neuronales Fuente: Propia*

Si a estas previsiones de la tendencia se añaden los datos de la estacionalidad, se obtienen las previsiones de las unidades de venta, representadas en la Figura 6.29 y en la Figura 6.30. El error relativo de estas previsiones es del 19,56 %. Cabe destacar que si se refina la etapa de diseño de la red neuronal se podrían obtener resultados más precisos; de todas formas, esto no se considera oportuno debido al tiempo necesario para realizar dicho diseño respecto al empleado para el resto de modelos candidatos.



*Figura 6.29 Previsiones de ventas del departamento 1 al aplicar modelo de redes neuronales Fuente: Propia*

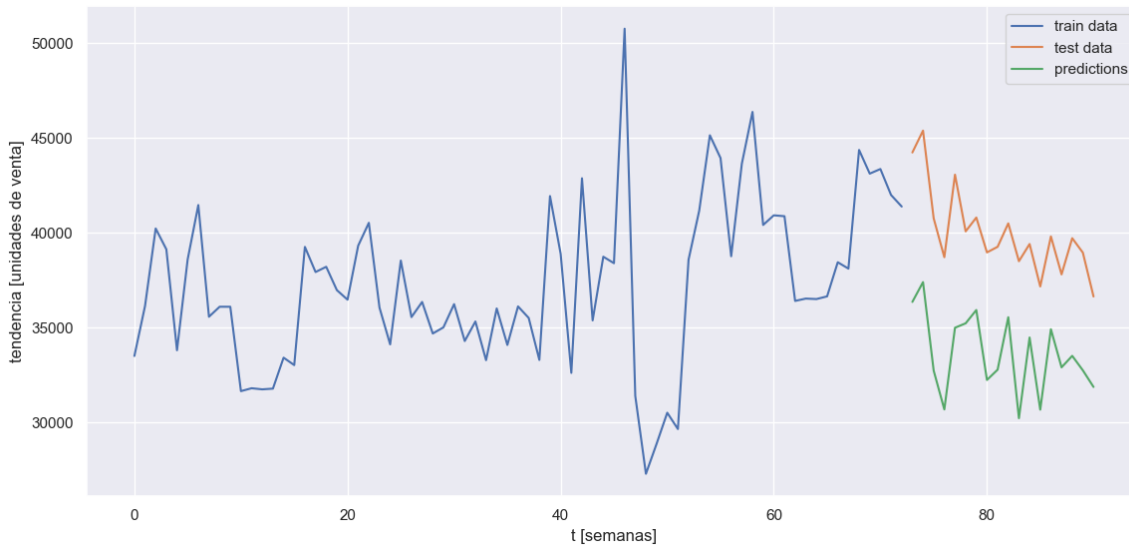


Figura 6.30 Previsiones de ventas del departamento 10 al aplicar modelo de redes neuronales

Fuente: Propia

### 6.5.7. Modelo XGBoost

El último modelo considerado es el basado en el algoritmo de conjunto *XGBoost*. Con dicho algoritmo también se predice la tendencia de las ventas y se suman los datos de la estacionalidad con el fin de conocer las previsiones de venta. El algoritmo utilizado para construir este modelo es el *XGBRegressor* de la librería *xgb* de Python [40].

Con el fin de conocer los valores más adecuados de los parámetros para construir el modelo, se realiza de nuevo una búsqueda de cuadrículas. Algunos de los resultados obtenidos de dicho análisis son que se deben considerar un total de 2000 árboles de decisión, que cada uno de ellos debe tener una profundidad máxima de 6 nodos y que en cada árbol de decisión se deben considerar un 50% de los datos [41].

Se construye un único modelo *XGBoost* para la totalidad de los datos, sin diferenciar por departamento, ya que este es capaz de realizar previsiones adecuadas considerando el atributo departamento como parte del modelo; convirtiéndose este en el atributo de mayor importancia a la hora de realizar las previsiones, tal y como puede observarse en la Figura 6.31.

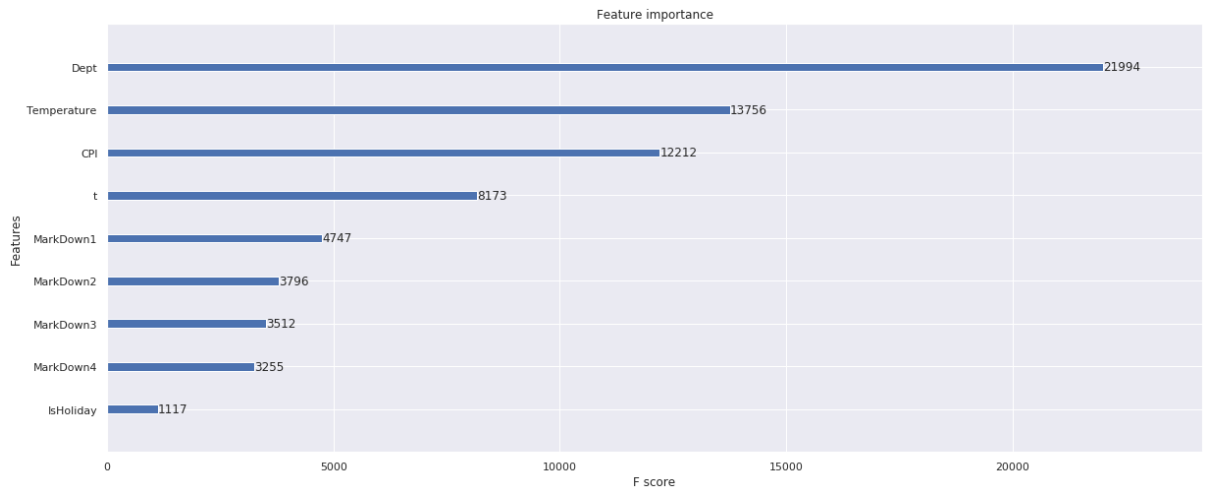


Figura 6.31 Importancia de los diferentes atributos X en el modelo XGBoost

Fuente: Propia

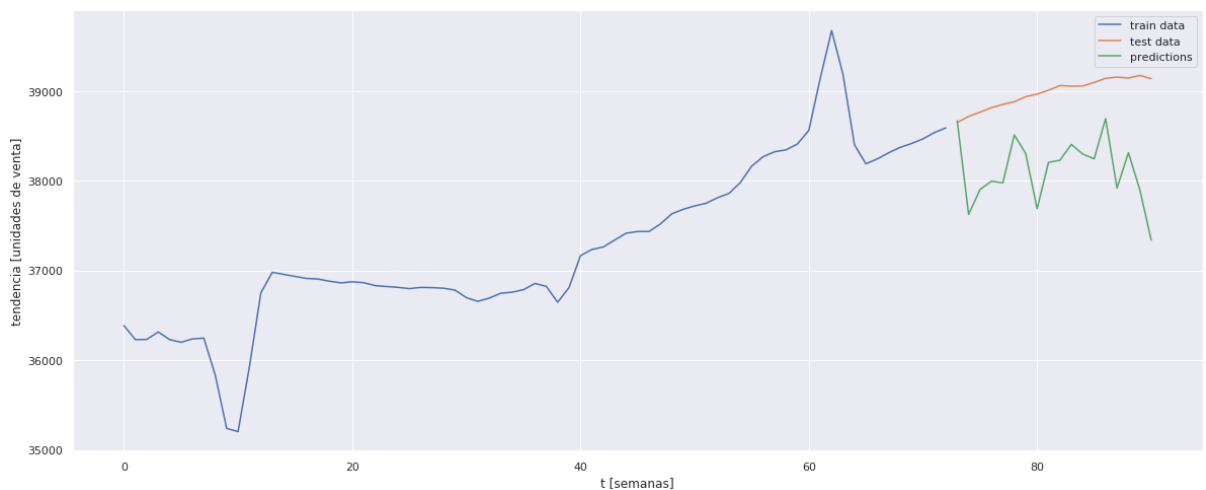
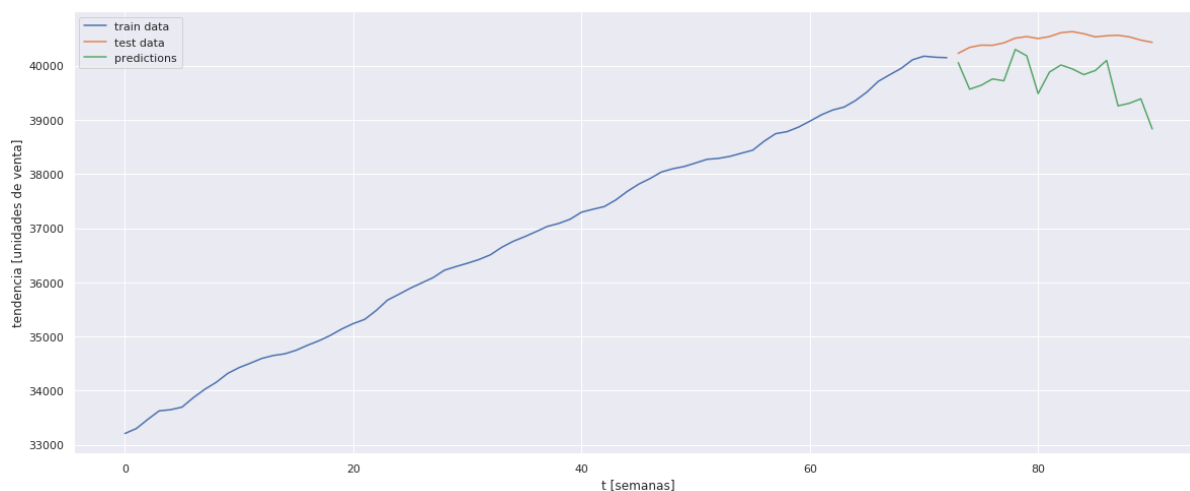


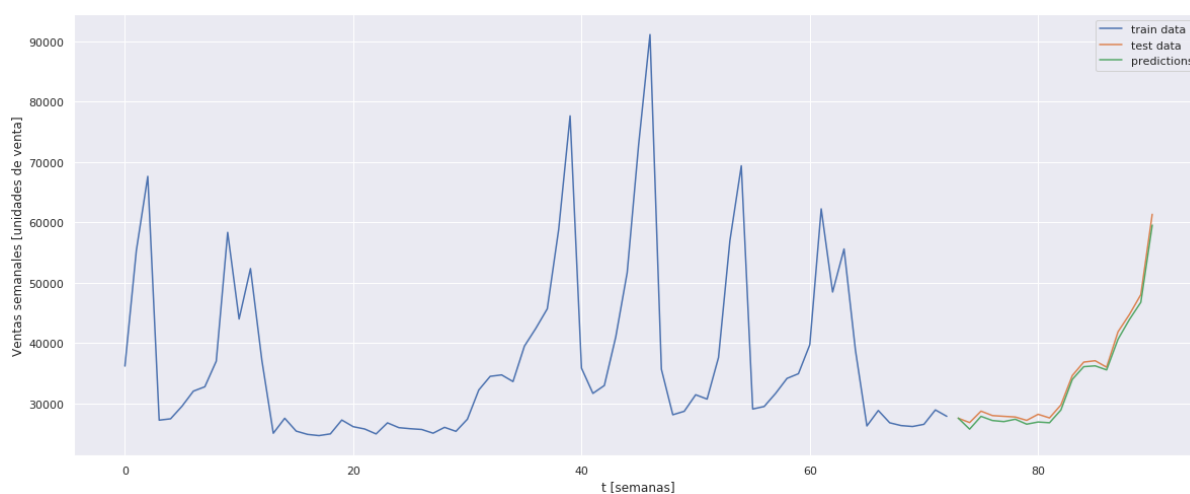
Figura 6.32 Previsiones de la tendencia de las ventas del departamento 1 al aplicar el modelo XGBoost

Fuente: Propia



*Figura 6.33 Previsiones de la tendencia de las ventas del departamento 10 al aplicar el modelo XGBoost Fuente: Propia*

En la Figura 6.32 y en la Figura 6.33 se representan las previsiones de la tendencia y a continuación, las previsiones de las unidades de venta semanales de los departamentos uno y diez. El error relativo global del modelo *XGBoost* al realizar las previsiones de la ventas dadas las entradas de los datos de prueba es del 16,88 %.



*Figura 6.34 Previsiones de ventas del departamento 1 al aplicar el modelo XGBoost Fuente: Propia*

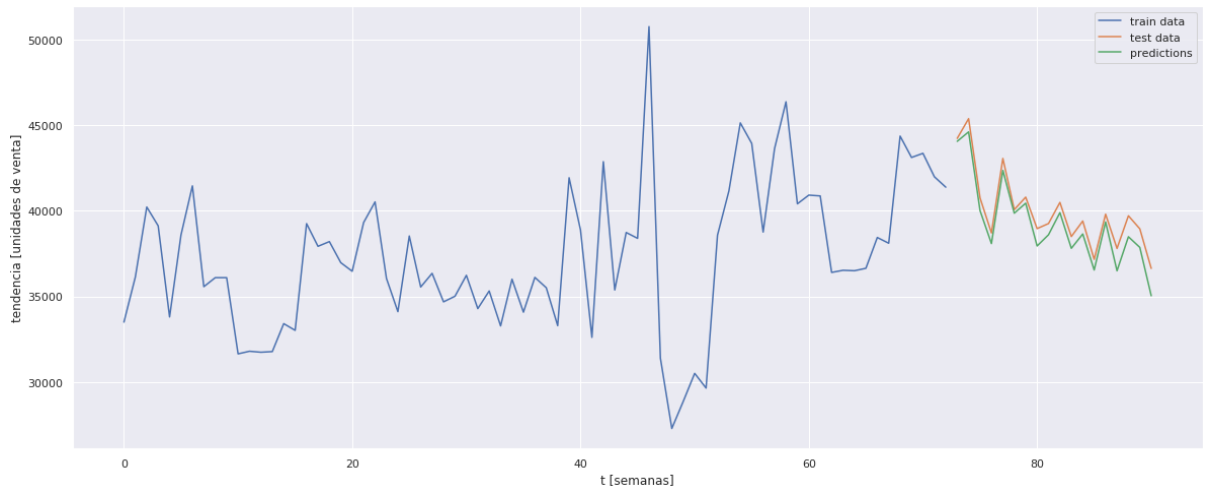


Figura 6.35 Previsiones de ventas del departamento 10 al aplicar el modelo XGBoost

Fuente: Propia

## 6.6. Selección del modelo definitivo

En el presente apartado se comparan los modelos candidatos presentados y se propone un modelo definitivo.

### 6.6.1. Comparación de los modelos candidatos considerados

En la Tabla 6.1 se recogen los errores relativos y las precisiones de los diversos modelos candidatos considerados. Estas métricas evalúan el rendimiento del modelo para predecir las unidades de venta dadas las entradas de los datos de prueba. La precisión se calcula como la diferencia entre la unidad y el error relativo.

Modelo candidato	Error relativo	Precisión
SMA	114,85 %	-
SARIMAX	31,29 %	68,71 %
Regresión lineal	7,06 %	92,94 %
k-vecinos más cercanos	16,85 %	83,15 %
Bosques aleatorios	3,20 %	96,8 %
Redes neuronales	19,56 %	80,44 %

XGBoost	16,88 %	83,12 %
---------	---------	---------

Tabla 6.1 Error relativo y precisiones de los modelos candidatos Fuente: Propia

Como puede observarse, los modelos de aprendizaje automático considerados predicen las ventas de manera más precisa que los modelos de análisis de series temporales contemplados. Por otro lado, tanto el modelo de regresión lineal como el modelo de bosques aleatorios presentan un rendimiento muy elevado, prediciendo las ventas con una precisión superior al 90%. Por tanto, se podría considerar cualquiera de estos modelos como un buen candidato a modelo definitivo, ya que esta es la precisión que se deseaba alcanzar. Sin embargo, considerar previsiones que sean combinación de las dadas por más de un modelo permite aumentar la fiabilidad y reducir la varianza dadas entradas desconocidas. Por tanto, a continuación, se propone un modelo definitivo en base al método de aprendizaje automático.

### 6.6.2. Selección del modelo definitivo en base al método de aprendizaje combinado

Se decide seleccionar como modelo definitivo un modelo basado en aquellos dos modelos candidatos que presentan una precisión mayor y superior al 90%: el modelo de bosques aleatorios y el de regresión lineal. Debido a que el primero de ellos presenta una precisión superior se considera adecuado seleccionar el modelo utilizando el método de media ponderada, con el fin de considerar en mayor medida las previsiones realizadas por el modelo más preciso. Tras comprobar diversas variaciones de dichas ponderaciones, se selecciona como modelo definitivo:

$$previsiones = 0,7 \cdot previsiones_{\text{Bosques aleatorios}} + 0,3 \cdot previsiones_{\text{Regresión lineal}} \quad (\text{Ec. 6.2})$$

Es decir, las previsiones de las ventas consideradas serán 0,7 del valor previsto por el modelo de bosques aleatorios más 0,3 del valor previsto por el modelo de regresión lineal. El modelo definitivo presenta un error del 3,43 % y, por tanto, una precisión del 96,57 %. A pesar de que dicha precisión sea prácticamente igual pero ligeramente inferior a la que proporciona el modelo de bosques aleatorios considerando los datos de prueba; se estima que, contemplando un banco de pruebas mayor, la precisión del modelo de aprendizaje combinado sería probablemente superior ya que las previsiones tenderían a presentar menos varianza.

En la Figura 6.36 y en la Figura 6.37 se representan las previsiones de las ventas semanales realizadas por el modelo definitivo dadas las entradas de los datos de prueba para los departamentos uno y diez. La inspección visual evidencia la alta precisión de dichas

previsiones. El modelo definitivo es un único modelo para todos los departamentos, ya que es suficiente para alcanzar la precisión deseada; aun así, considera diferentes parámetros para las diferentes secciones del establecimiento.

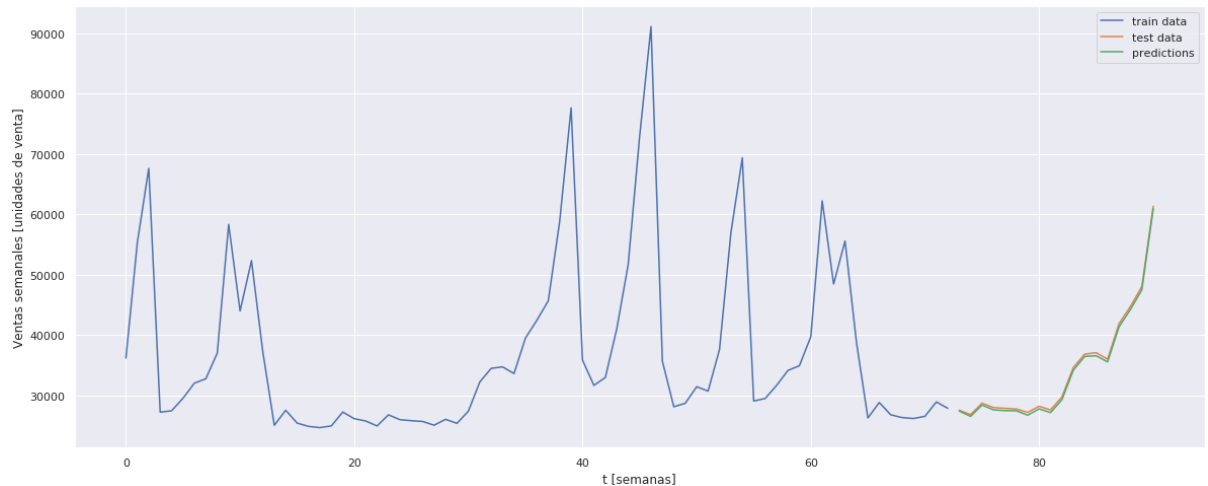


Figura 6.36 Previsiones de ventas del departamento 1 al aplicar el modelo definitivo

Fuente: Propia

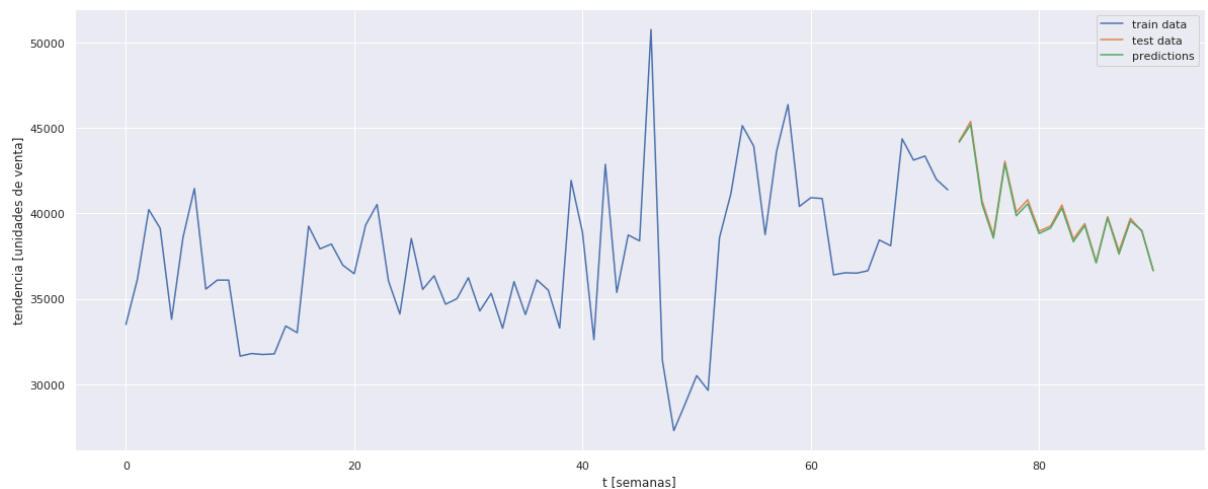


Figura 6.37 Previsiones de ventas del departamento 10 al aplicar el modelo definitivo

Fuente: Propia

## 6.7. Aplicación y actualización del modelo definitivo

El modelo definitivo presentado permite predecir las ventas de los 98 departamentos del establecimiento Walmart con una precisión del 96,57 % considerando los datos de prueba. Dicha precisión es la esperada a la hora de realizar previsiones dadas nuevas entradas, ya que los datos de prueba no han sido utilizados para entrenar el modelo de aprendizaje

automático definitivo.

Tal y como ha sido expuesto con anterioridad, en caso de disponer de los datos históricos de las unidades vendidas de cada uno de los productos, se debería realizar un modelo de previsión que permitiera predecir las ventas de cada familia por separado. La precisión que ha sido obtenida en el modelo definitivo sería adecuada para aquellos productos considerados críticos, según el criterio presentado en la sección 6.2. Por otro lado, una precisión menor y alrededor del 80%, la cual ha sido alcanzada por todos los modelos candidatos de aprendizaje automático considerados, sería suficiente para las familias no críticas. Por ello, a la hora de establecer un modelo de previsión de ventas para los productos no críticos no sería necesario invertir tantos esfuerzos en la etapa de diseño, ya que considerando un único modelo candidato de aprendizaje automático sería probablemente suficiente.

A la hora de utilizar el modelo de previsión de ventas diseñado, es importante valorarlo regularmente a nivel cualitativo. En el caso de Walmart, los resultados numéricos deben ser juzgados por líderes y expertos de ventas de la compañía. Las previsiones serán realizadas por un departamento técnico de operaciones, pero es de vital importancia la comunicación entre dicho departamento y el departamento de ventas y *marketing* con el fin de obtener unas previsiones fiables. Por un lado, las decisiones estratégicas que sean adoptadas por ventas y/o *marketing* deben ser incluidas como datos a considerar en el modelo; tal y como se ha hecho en la resolución del caso práctico al considerar las acciones comerciales (*Markdowns*). Por otro lado, el departamento de ventas será el responsable de juzgar la adecuación de las previsiones numéricas obtenidas y de detectar posibles anomalías en los resultados; como, por ejemplo, ventas muy superiores o inferiores a las esperadas y no causadas por un acontecimiento extraordinario conocido. La detección de anomalías continuadas implicará una necesidad de revisión y actualización del modelo de previsión.

Como se ha visto con anterioridad, una previsión de ventas precisa permite evitar fluctuaciones en los niveles de existencias e ineficiencias en la gestión de actividades de los detallistas, en este caso del establecimiento Walmart, y de todos los actores involucrados en la cadena de suministro. A continuación, se destacan algunas ventajas concretas de aplicar el modelo de ventas diseñado en Walmart. También se comentan algunas de las barreras que se pueden encontrar a la hora de implementar satisfactoriamente dicho modelo.

#### **6.7.1. Ventajas de aplicar el modelo de previsión de ventas**

Aplicar un modelo de previsión de ventas con una precisión como el diseñado, permite aumentar la satisfacción de los clientes a la vez que mejorar la gestión de los procesos de la



cadena de suministro.

Predecir las ventas, en caso de que estas reflejen correctamente la previsión de la demanda, permite aumentar la satisfacción de los clientes. Conocer que desearán consumir los clientes permite a los establecimientos disponer del producto necesario en el momento necesario y mejorar así la relación con el consumidor. Este punto se considera de gran relevancia ya que, si la satisfacción del consumidor crece, estos consumirán más y los ingresos de la empresa aumentarán.

Se debe destacar también que los modelos de previsión de ventas se pueden utilizar para analizar los efectos que puede generar el cambio de precios o el impacto de estrategias comerciales; es decir, se pueden utilizar como herramientas para tomar decisiones estratégicas.

Por otro lado, conocer con anterioridad que se venderá, permite a los establecimientos mantener unos niveles de existencias estables y bajos. Predecir correctamente las ventas permitirá a los establecimientos aplicar el método de justo a tiempo (*Just In Time*) y disponer del stock necesario reduciendo en gran medida los costes logísticos. En una empresa como Walmart, donde principalmente se comercializan productos de consumo de alta rotación o *fast-moving consumer goods* en inglés, los márgenes de beneficio son bajos y un gran porcentaje de los costes totales son debidos a las existencias. Por tanto, es importante implantar herramientas o procesos que permitan reducir los costes asociados al inventario.

Se debe considerar también que un gran porcentaje de los productos comercializados disponen de fecha de caducidad, una previsión de ventas adecuada permitirá reducir el riesgo de disponer de stock obsoleto de productos con una frescura inadecuada. Además, el modelo de previsión permitirá predecir la demanda de nuevos productos; por lo que se podrán reducir también los elevados niveles de stock de seguridad que se suelen considerar para nuevos lanzamientos.

Disponer de unas previsiones adecuadas elimina en gran medida la necesidad de realizar acciones o procesos no planeados con el fin de satisfacer demanda inesperada. En el caso del establecimiento Walmart, los departamentos comerciales presionarían a operaciones para conseguir más cantidad de producto si observan que un artículo tiene más éxito del esperado. Con el fin de aumentar la satisfacción del cliente y los ingresos de la compañía, operaciones necesitaría realizar esfuerzos extras para abastecerse de producto y satisfacer la demanda inesperada. Considerando la filosofía y metodología *Lean* de gestión de procesos, la necesidad de realizar estas tareas es una pérdida (*waste*) ya que no añade valor al proceso desde el punto de vista del consumidor; y, por tanto, dicha situación debe ser evitada.

Se debe destacar también que si el establecimiento minorista, en este caso Walmart, dispone

de unas previsiones de venta adecuadas, los proveedores de los diferentes productos pueden planificar las actividades de producción y transporte sin considerar un número elevado y continuo de imprevistos. Esto permitirá establecer una relación *win-win* con los proveedores, relación deseada ya que permitirá a ambos actores crecer y evolucionar gracias a la relación de beneficio conjunto.

### **6.7.2. Posibles barreras que pueden impedir la correcta implementación del modelo de previsión de ventas**

A la hora de aplicar el modelo podría observarse que la precisión de las previsiones es inferior a la esperada. Esto podría deberse a diferentes causas y, de ser detectadas, el modelo debería actualizarse acordemente.

Por un lado, el modelo de previsión de ventas se basa en las ventas pasadas y la demanda puede que haya sido superior a la capacidad de venta del establecimiento Walmart. Si a la hora de aplicar el modelo de previsión se observase que las ventas siempre son superiores a las esperadas, o las máximas que el establecimiento es capaz de abastecer, el modelo no estaría prediciendo correctamente la demanda y debería ser actualizado. Con el fin de evitar este inconveniente, es preciso registrar e incorporar como variable del modelo si se está vendiendo a máxima capacidad.

Por otro lado, podría detectarse que existe exceso de stock para unos productos y carencia de este para otros productos de la misma familia. En caso de observarse este fenómeno sería necesario revisar la agrupación de los artículos y separar ambos productos en familias diferentes. Para poder detectar tanto esta casuística como otras, es importante disponer de una recogida de datos fiable y que estos se recojan por producto; a pesar de que las previsiones se hagan por grupos de productos.

Se debe destacar también la importancia de disponer de una comunicación constante con los departamentos de ventas y *marketing* con el fin de que los acontecimientos futuros relacionados con los cambios de precio o estrategias comerciales sean correctamente reflejados en los datos utilizados por los modelos de previsiones de venta. Dicha comunicación será también necesaria para detectar si el modelo realiza previsiones anómalas y necesita, por tanto, ser actualizado.

## 7. Planificación temporal y costes

En la planificación temporal del proyecto, presentada en la Tabla 7.1, se incluyen todas las actividades, y la duración de estas, realizadas para confeccionar el presente trabajo. En primer lugar, se realizó un curso online de aprendizaje automático en la plataforma *coursera*, titulado *Machine Learning by Stanford University* e impartido por el profesor Andrew Ng. Durante este período de 11 semanas, también se realizaron actividades de formación complementarias, la mayoría de ellas participando en tertulias y actividades realizadas por grupos de interés en la empresa Amazon, organización donde trabaja la autora del proyecto. Al final de este período, se seleccionó el caso práctico previamente presentado.

A continuación, se procedió a estudiar y analizar las diversas técnicas de previsión de ventas existentes y entender tanto la fundamentación matemática de dichas técnicas como la posible aplicación de estas en el caso práctico; además de las funciones de Python que necesitaban ser utilizadas para ello. Durante este período y durante las primeras semanas en que se trabajó en el caso práctico, se analizó la metodología presentada para la selección de un modelo de previsión. Una vez ésta definida, se trabajó en aplicar diversas técnicas de previsión dados los datos del caso práctico y se obtuvo, tras seis semanas, el modelo definitivo. Por último, se valoró el impacto de implementar el resultado del caso realizado y se confeccionó la memoria, completando las 31 semanas que todo el proceso del Trabajo de Fin de Máster ha conllevado.

Tarea	Semanas	S1	...	S31
Curso online de aprendizaje automático	11			
Selección del caso práctico	2			
Estudio y análisis de diversas técnicas de previsión de ventas	4			
Caso práctico - análisis y manipulación de los datos	2			
Caso práctico - aplicación de las diversas técnicas de previsión	6			
Redacción de la memoria	8			

Tabla 7.1 Planificación temporal del proyecto

Fuente: Propia

Por otro lado, en la Tabla 7.2, se presentan los costes de la realización del proyecto. Para el cálculo del coste total se ha tenido presente el uso del material de oficina y de los soportes informáticos empleados y se ha estimado el salario asociado al desarrollo del proyecto. El salario ha sido calculado en base al salario medio de los científicos de datos, profesión relacionada con el presente proyecto, en el país de Luxemburgo [42], país en que dicho trabajo ha sido confeccionado.

Concepto	Precio unitario	Unidades	Total
Material de oficina			40.00 €
Licencias sistemas informáticos	1,150.00 €	1	1,150.00 €
Salario <i>Data Scientist</i>	25 €/h	360 h	9,000.00 €
			<b>10,190.00 €</b>

Tabla 7.2 Costes del proyecto Fuente: Propia

## 8. Impacto ambiental

El efecto ambiental que implica el desarrollo del presente proyecto es mínimo ya que se trata de un estudio analítico y los recursos energéticos empleados en el transcurso del mismo son despreciables. De todas formas, la implementación de los resultados del caso práctico realizado sí comportaría un impacto ambiental.

Si se aplica el modelo de ventas diseñado para el establecimiento Walmart, tal y como se ha analizado en la sección 6.7, la eficiencia de la cadena de suministro de la compañía aumentaría. Además, la mejora de la planificación de algunos procesos de la cadena permitiría disminuir el impacto ambiental de la compañía.

En concreto, se podría disminuir notablemente la cantidad de producto obsoleto por pérdida de frescura. Walmart comercializa un gran número de artículos alimenticios que disponen de fecha de caducidad. Si no se predice correctamente las cantidades vendidas de dichos productos, la compañía se ve obligada a deshacerse de un gran número de artículos caducados. Alrededor del 20% de estos productos son donados a ONGs; el resto de productos son destruidos, con el impacto ambiental que ello conlleva. Actualmente cada año se generan 1600 millones de toneladas de residuos alimenticios debido a productos caducados y la *UN's Sustainable Development Goals* establece que en 2030 los residuos deberían reducirse a la mitad [43]. Por tanto, implementar medidas que permitan la reducción de estos residuos resulta imprescindible para alcanzar los objetivos ambientales establecidos por las Naciones Unidas.

Además, el modelo de previsión de ventas considerado permite también predecir la demanda de productos nuevos. En general, se consideran stocks de seguridad muy elevados para los nuevos lanzamientos, provocando esto un número de obsoletos superiores al generado por el resto de artículos. Al igual que para los productos con una frescura inadecuada, los stocks obsoletos de nuevos productos deben ser también donados o destruidos generando esto un impacto ambiental indeseado.

Por tanto, se concluye que la creación de un modelo de previsión de ventas preciso permite la reducción de stocks obsoletos e impide la destrucción de estos evitando el consecuente impacto ambiental asociado.



## Conclusiones

Del análisis del impacto de las previsiones de ventas en la gestión de la cadena de suministro se concluye que una previsión adecuada es necesaria para realizar una gestión eficiente de todas las actividades asociadas a la cadena de abastecimiento. Previsiones poco precisas provocan que las cantidades vendidas sean inesperadas y ello conlleva fluctuaciones en las existencias de los diferentes actores de la cadena generando esto una mala gestión de los recursos disponibles. Existen también casos en los que disponer de unas previsiones de venta precisas es más crítico que en otros; en concreto, una precisión elevada es necesaria para los productos que generan un gran número de ventas y/o ingresos, que tienen un elevado coste de stock, un alto riesgo a generar stock obsoleto o un tiempo de abastecimiento alto.

Del análisis de las técnicas y métodos de previsión de ventas realizado, se extrae que existen dos enfoques para realizar previsiones: realizarlas con métodos cualitativos o con métodos cuantitativos. En el caso de las metodologías cualitativas es indispensable la intuición y experiencia de expertos para aplicarlas a casuísticas concretas. De lo contrario, las cuantitativas se basan en técnicas estadísticas o herramientas matemáticas que pueden adaptarse a casos particulares sin requerir conocimiento explícito para cada caso particular. Se considera la importancia de utilizar técnicas cuantitativas para realizar previsiones de venta y valorar los resultados mediante criterios cualitativos.

Se concluye también que existen dos enfoques principales dentro de los métodos cuantitativos utilizados para realizar previsiones de ventas, y que pueden ser aplicados a cualquier tipo de previsión; estos son: los métodos basados en el análisis de series temporales y los métodos de causalidad. Los primeros detectan la tendencia de los datos y la proyectan para obtener las previsiones; los segundos, generalmente más precisos, determinan relaciones causa-efecto y a partir de ellas generan las previsiones.

Debido a la gran diversidad de enfoques y métodos de previsión existentes y las diferentes necesidades que presenta cada casuística, se detecta la necesidad de presentar una metodología que permita, dado un caso concreto, realizar un modelo de previsión adecuado. Con el fin de obtener un modelo con la precisión estimada oportuna, se plantea la necesidad de contemplar diversos modelos candidatos y realizar una valoración de qué modelo o combinación de ellos permite unas mejores previsiones. En consecuencia, se considera oportuno dividir el histórico de datos en datos de entrenamiento y datos de prueba; con el fin de utilizar los primeros para entrenar los algoritmos del modelo cuantitativo, en caso de que esto sea necesario, y los datos de prueba para valorar la precisión de las previsiones que cada modelo candidato proporciona. Se destaca también la necesidad de escoger los valores de los parámetros de los algoritmos en que se basa cada uno de los modelos en función de aquellos valores que permitan unas mejores previsiones; para ello se proponen dos

metodologías: la búsqueda aleatoria y la búsqueda de cuadrículas.

En el caso práctico presentado se desea predecir las ventas de un establecimiento, donde la mayoría de productos comercializados son artículos de consumo de alta rotación. Con el fin de definir un modelo de previsión de ventas que permita aumentar la eficiencia de la compañía sin recurrir en grandes costes de diseño de este, se considera la necesidad de realizar las previsiones por grupos de familia. Además, se establece que un 20% se deben considerar familias críticas, ya que serán las que causarán un mayor impacto en la gestión de la cadena de suministro, y la previsión de sus ventas debe ser precisa. Entre las familias críticas se debe considerar las que generar un gran número de ventas y/o ingresos, las que tienen un tiempo de abastecimiento alto, las nuevas familias y aquellas donde la caducidad de sus productos, de existir, es crítica.

Debido a que no se dispone de los datos de venta por producto y familia, si no por departamento, se decide realizar las previsiones de ventas por departamento y diseñar un modelo de alta precisión aplicable a las familias críticas. Del análisis de los datos se extrae la posibilidad de o bien utilizar métodos basados en el análisis de series de tiempo con el fin de detectar la tendencia y estacionalidad y proyectarla para realizar previsiones; o bien, utilizar modelos causales de aprendizaje automático para predecir la tendencia de los datos y considerar la estacionalidad como un dato conocido para cada semana del año.

De los resultados de aplicar la metodología presentada en el caso práctico, se observa que los modelos de previsión basados en el análisis de series temporales presentan un mayor error y menor precisión que los modelos de aprendizaje automático considerados. En concreto, el modelo de series temporales que predice con menor error las previsiones es el modelo SARIMAX, el cual presenta una precisión del 68,71 %. Por otro lado, el modelo de aprendizaje automático con mayor error es el basado en redes neuronales, con una precisión del 80,44 %. De todos los modelos considerados, aquellos que presentan unos mejores resultados son el modelo basado en el algoritmo de bosques aleatorios y el modelo de regresión lineal; con unas precisiones respectivas del 96,8 % y del 92,94 %.

Con el fin de proponer un modelo robusto y con poca varianza para realizar las previsiones de venta, se selecciona como modelo definitivo uno basado en el método de aprendizaje combinado. Se propone predecir las ventas en base a las previsiones realizadas por los dos modelos candidatos más precisos y considerando en mayor medida el más preciso de ambos. Por tanto, el modelo definitivo propuesto considera las previsiones del modelo de bosques aleatorios, en un 70%, y las del modelo de regresión lineal, en un 30%. Dicho modelo definitivo predice las ventas dadas las entradas de los datos de prueba con una precisión del 96,57 %.



Esta precisión puede considerarse como la precisión esperada dadas entradas desconocidas, ya que los datos de prueba no han sido utilizados para entrenar al modelo.

De la valoración de aplicar el modelo de previsión de ventas en la compañía Walmart, se destaca que permitiría aumentar la satisfacción del cliente a la vez que aplicar el método de gestión de inventario *Just in Time*. Por tanto, se reducirían los costes de stock y disminuirían los niveles de stock obsoleto evitando la destrucción de estos. Las previsiones de venta repercutirían también, de manera positiva, en la gestión de actividades e inventarios del resto de actores de la cadena de suministro; y, por tanto, permitiría a Walmart establecer una relación *win-win* con sus proveedores.

Como futuras líneas de trabajo se destaca la posibilidad de establecer un modelo automático de diseño de modelos de previsión de ventas que permita adaptarse, a partir de ciertos *inputs*, a casuísticas concretas. Respecto al caso práctico, se propone minimizar el riesgo de que las posibles barreras a la hora de implementar el modelo de previsiones diseñado en Walmart, impidan el correcto funcionamiento de este. Para ello, se destaca la necesidad de trabajar en recoger los datos de las unidades vendidas por producto y de realizar un análisis *clustering* para la correcta división de los productos en familias.



## Agradecimientos

Doy las gracias a mis compañeros de trabajo de Amazon, por entusiasmarme por la Industria 4.0, la ciencia de datos y el *Machine Learning*, conceptos en los que se basa este proyecto. Gracias a todos ellos y en especial a Andrei, por su guía en la fase de aprendizaje de este trabajo.

Gracias también a mi mentora dentro de la compañía, Rachel Cooke, por su constante insistencia y motivación en todas las fases del proyecto con el fin de que éste se convirtiera en una realidad.

Gracias también a mi familia y amigos, por lidiar con mis frustraciones y alegrías durante el desarrollo del proyecto.



## Bibliografía

- [1] N. Jazdi, «Cyber physical systems in the context of Industry 4.0,» *IEEE*.
- [2] J. H. C. SOTERO, «Logistweb,» 9 Septiembre 2008. [En línea]. Available: <https://logistweb.wordpress.com/2008/09/09/actores-en-la-cadena-de-abastecimiento-scm/>.
- [3] W. D. J. S. K. S. M. N. W. N. C. D. S. Z. G. Z. John T. Mentzer, «DEFINING SUPPLY CHAIN MANAGEMENT,» *Journal of Business Logistics*, vol. 22, nº 2, 2011.
- [4] H. L. Lee, V. Padmanabhan y S. Whang, «The Bullwhip Effect in Supply Chains,» *Sloan Management Review*, 1997.
- [5] Departament d'Organització d'Empreses ETS d'Enginyeria Industrial de Barcelona, Diseño de la cadena de suministro - Previsiones, Barcelona: Universitat Politècnica de Catalunya, 2018.
- [6] M. C. I. N. L. Rodríguez y G. F. Preciado, Series de tiempo, Hermosillo: Universidad de Sonora.
- [7] M. Arellano, «Introducción al Análisis Clásico de Series de Tiempo,» 5campus.com, Estadística, 2001. [En línea]. Available: <http://www.5campus.com/leccion/seriest>.
- [8] «devtodev,» 09 Agosto 2017. [En línea]. Available: <https://edu.devtodev.com/articles/83/seasonality-of-the-project-do-not-be-afraid-of-summer-recession>.
- [9] R. J. Hyndman y G. Athanasopoulos, «Autoregressive models,» de *Forecasting: Principles and Practice*, otexts, 2018.
- [10] R. Nau, «The mathematical structure of ARIMA,» Fuqua School of Business, Duke University, Durham, 2014.
- [11] E. H. Etuk, «An Additive SARIMA Model for Daily Exchange Rates of the,» *International Journal of Empirical Finance*, vol. 2, nº 4, pp. 193-201, 2014.
- [12] M. Hassan, M. Islam, M. Imam y S. Sayem, «Forecasting wholesale price of coarse

rice in Bangladesh: A seasonal autoregressive integrated moving average approach,» *Journal of the Bangladesh Agricultural University*, vol. 11, nº 2, pp. 271-276, 2013.

- [13] Amazon EU Confidential, «LSTM for Time Series Forecasting,» 2019.
- [14] J. Mammadov, «Econometric Approach to Time Series Analysis — Seasonal ARIMA in Python,» *Towards Data Science*, 16 Agosto 2019.
- [15] A. Buteikis, Time series with trend and seasonality, Vilna: Universidad de Vilna.
- [16] A. Ng, «Machine Learning online course offered by Stanford,» coursera.
- [17] J. Brownlee, «Machine Learning Mastery,» [En línea]. Available: <https://machinelearningmastery.com/>.
- [18] A. Tiwari, «Understanding Overfitting and Underfitting in Machine Learning,» *Towards Data Science*, 6 Septiembre 2019.
- [19] M. Stewart, «The Actual Difference Between Statistics and Machine Learning,» *Towards Data Science*, 25 Marzo 2019.
- [20] R. Gandhi, «Introduction to Machine Learning Algorithms: Linear Regression,» *Towards Data Science*, 27 Mayo 2018.
- [21] B. Boehmke y B. Greenwell, «Chapter 7 Multivariate Adaptive Regression Splines,» de *Hands-On Machine Learning with R*, Taylor & Francis Group, 2019.
- [22] J. P. Figueira, «LOESS Smoothing data using local regression,» *Towards Data Science*, 24 Mayo 2019.
- [23] S. Shaier, «ML Algorithms: One SD ( $\sigma$ )- Instance-based Algorithms,» *Towards Data Science*, 2 Febrero 2019.
- [24] A. SINGH, «Analytucs Vidhya,» 22 Agosto 2018. [En línea]. Available: <https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/>. [Último acceso: Noviembre 2019].
- [25] D. S. Sayad, «An Introduction to Data Science - K Nearest Neighbors Regression,» [saedsayad.com](http://saedsayad.com), 2019.

- [26] J. O. Alvear, «4 Árboles de Decisión - Parte II,» de *Árboles de decisión y Random Forest*, Cuenca, bookdown.org, 2018.
- [27] N. S. Chauhan, «Decision Tree Algorithm — Explained,» *Towards Data Science*, 24 Diciembre 2019.
- [28] J. Schmidhuber, «Deep Learning in neural networks: An overview,» *ELSEVIER*, 2014.
- [29] T. G. Dietterich, «Ensemble Learning,» *MIT Press*, 2002.
- [30] E. Lee, «An Intro to Hyper-parameter Optimization using Grid Search and Random Search,» *Towards Data Science*, 5 Junio 2019.
- [31] pandas developers, «pandas.DataFrame.rolling,» [En línea]. Available: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html>.
- [32] J. Salvi, «Significance of ACF and PACF Plots In Time Series Analysis,» *Towards Data Science*, 27 Marzo 2019.
- [33] Josef Perktold, Skipper Seabold, Jonathan Taylor, statsmodels-developers, «statsmodels.tsa.statespace.sarimax.SARIMAX,» [En línea]. Available: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>. [Último acceso: Noviembre 2019].
- [34] scikit-learn developers, «sklearn.linear\_model.LinearRegression,» scikit learn, [En línea]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LinearRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html). [Último acceso: Noviembre 2019].
- [35] scikit-learn developers, «sklearn.preprocessing.StandardScaler,» [En línea]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>. [Último acceso: Noviembre 2019].
- [36] scikit-learn developers, «sklearn.neighbors.KNeighborsRegressor,» [En línea]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html>.

[Último acceso: Noviembre 2019].

[37] scikit-learn developers, «sklearn.ensemble.RandomForestRegressor,» [En línea]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>. [Último acceso: Noviembre 2019].

[38] keras developers, «Keras Documentation,» [En línea]. Available: <https://keras.io/models/sequential/>. [Último acceso: Noviembre 2019].

[39] keras developers, «Keras Optimizers,» [En línea]. Available: <https://keras.io/optimizers/>. [Último acceso: Noviembre 2019].

[40] xgboost developers, «XGBoost Python Package,» [En línea]. Available: <https://xgboost.readthedocs.io/en/latest/python/index.html>. [Último acceso: Noviembre 2019].

[41] xgboost developers, «XGBoost Parameters,» [En línea]. Available: <https://xgboost.readthedocs.io/en/latest/parameter.html>.

[42] «Average Data Scientist Salary in Luxembourg,» PayScale, [En línea]. Available: [https://www.payscale.com/research/LU/Job=Data\\_Scientist/Salary/20027d5b/Luxembourg](https://www.payscale.com/research/LU/Job=Data_Scientist/Salary/20027d5b/Luxembourg). [Último acceso: Diciembre 2019].

[43] E. HEGNSHOLT, S. UNNIKRISHNAN, M. POLLMANN-LARSEN, B. ASKELSDOTTIR y M. GERARD, «BCG - Tackling the 1.6-Billion-Ton Food Loss and Waste Crisis,» [En línea]. Available: <https://www.bcg.com/publications/2018/tackling-1.6-billion-ton-food-loss-and-waste-crisis.aspx>. [Último acceso: Enero 2020].